Behavioral/Cognitive

Neural Distinction between Visual Word and Object Recognition: An fMRI Study Using Pictographs

Jiahong Zeng,[1] Yudan Luo,[1,2] Xiangqi Luo,[1] Saiyi Jiao,[1,3] Ke Wang,[1,4] Zhenjiang Cui,[1] Chunyu Zhao,[1] Zhiyun Dai,[1] Yuxin Liu,[1] Yidong Jiang,[1] and Zaizhu Han[1]

[1]National Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China, [2]Department of Psychology and Art Education, Chengdu Education Research Institute, Chengdu 610036, China, [3]Department of Otorhinolaryngology, Peking Union Medical College Hospital, Beijing 100730, China, and [4]School of System Science, Beijing Normal University, Beijing 100875, China

It remains an open question in visual neuroscience whether the recognition of written words and visual objects engages distinct neural mechanisms intrinsically, unaffected by confounding factors such as stimulus properties and task demands, and, if so, where these differences are localized. Previous studies comparing these two processes have faced challenges in simultaneously controlling stimulus properties, including low-level visual features and high-level phonological and semantic attributes, as well as task demands. Here, we addressed these issues using Chinese pictographs, visually identical stimuli that can be interpreted either as words (lexical symbols) or as objects (visual depictions) and that were rigorously matched in a visual form, phonology, and semantics. During functional magnetic resonance imaging, 36 male and female human participants performed three language tasks (realness judgment, sound retrieval, and meaning judgment) on pictographs that were contextually recognized as words or objects, with each task applied to both recognition types under identical procedures. Results revealed robust word–object differences in the inferior parietal lobule (IPL), anterior cingulate cortex (ACC), and their associated networks. Compared with object recognition, word recognition elicited stronger activation in the IPL and reduced deactivation in the ACC. Furthermore, both regions exhibited distinct multivoxel activation patterns between the word and object recognition and showed stronger functional connectivity with other brain regions specifically during word recognition. This study provides well-controlled evidence for intrinsic neural dissociations between word and object recognition, highlighting a parietal–cingulate network as a core substrate differentiating these processes.

Key words: anterior cingulate cortex; inferior parietal lobule; object recognition; pictographs; word recognition

Significance Statement

Understanding how the brain distinguishes visual words from objects is fundamental to reading and visual cognition. However, previous studies have struggled to separate intrinsic neural differences from confounding factors like visual appearance, phonological and semantic content, and task demands. Using a novel design based on Chinese pictographs, visually identical stimuli interpretable as either words or objects, this study eliminates these confounds, enabling direct comparison under identical task. Functional magnetic resonance imaging results reveal robust differences in activation, multivariate pattern, and connectivity, highlighting the inferior parietal lobule, anterior cingulate cortex, and their associated networks as key neural substrates. These findings offer well-controlled evidence for intrinsic neural differences between word and object recognition, with implications for reading research, literacy education, and disorders like dyslexia.

Received Dec. 10, 2024; revised April 9, 2025; accepted May 13, 2025.

Author contributions: J.Z. and Z.H. designed research; J.Z., Y.L., and X.L. performed research; J.Z. and Z.C. analyzed data; J.Z., Y.L., S.J., K.W., Z.C., C.Z., Z.D., Y.L., Y.J., and Z.H. wrote the paper.

This work was supported by the Major Project of National Social Science Foundation (24&ZD252) and National Natural Science Foundation of China (32271091 and 82372555). We thank Mingyang Li and Chuqi Liu for their help with experimental design, data collection, data analysis, and visualization; Xiaojin Ma, Jianying Xu, and their colleagues for their suggestions on material visualization; Xinpeng Li for the help with paper writing; Yao Cheng, Wenli Liu, Meijuan Zhang, and Yixi Han for their supports provided during the research; the BNU-Han Lab members for their helpful suggestions on the paper; and all research participants for their cooperation.

The authors declare no competing financial interests.

Correspondence should be addressed to Zaizhu Han at zzhhan@bnu.edu.cn.

This paper contains supplemental material available at: https://doi.org/10.1523/JNEUROSCI.2322-24.2025

https://doi.org/10.1523/JNEUROSCI.2322-24.2025

Introduction

One of the central debates in visual neuroscience concerns whether the human brain processes written words differently from visual objects and, if so, where this differentiation occurs. While neuroimaging studies have identified distinct activations between words and objects in the ventral occipital–temporal cortex (e.g., visual word form area, VWFA; Cohen and Dehaene, 2004; Dehaene and Cohen, 2011; Saygin et al., 2016), the left angular gyrus (Menard et al., 1996; Seghier et al., 2010; Seghier, 2012), inferior parietal lobule (IPL; Devereux et al., 2013), inferior frontal gyrus (Vandenberghe et al., 1996; Shinkareva et al.,

2011), anterior cingulate cortex (ACC; Schlochtermeier et al., 2013), middle temporal gyrus (MTG; Köhler et al., 2000), and other regions (Bright et al., 2004; Shinkareva et al., 2011), a fundamental question remains: Are the observed neural differences driven by intrinsic recognition mechanisms, or do they arise from uncontrolled extraneous factors (e.g., stimulus physical properties and task demands)?

Traditional neural comparisons between written words and visual object recognition often encounter significant challenges. Early studies compared real words with real objects (Price et al., 2006; Shinkareva et al., 2011; Devereux et al., 2013; Caffarra et al., 2017; Giari et al., 2020), but this approach introduced fundamental differences in low-level visual and physical form between the two categories. These differences may not only influence early visual processing areas in the brain but also affect higher-level processes such as phonology and semantics through bottom–up mechanisms (Carreiras et al., 2014; Kay and Yeatman, 2017). To control for low-level visual processing difference, recent studies have trained participants to recognize identical meaningless shapes as either words or objects (Xue et al., 2006; Li et al., 2020; Y. Luo et al., 2024). However, these studies face challenges because the learned content differs, leading to differences in higher-level components (e.g., phonology and semantics; Saygin et al., 2016; Klaus and Hartwigsen, 2019; Kuhnke et al., 2023). These differences are reflected not only in higher-level brain areas but may also modulate form-related areas through top–down mechanisms (Carreiras et al., 2014). Beyond these confounding factors, task demands further complicate comparisons between word and object recognition (Pattamadilok et al., 2017). Overall, inherent differences across visual, phonological, and semantic dimensions, combined with uncontrolled task demands, confound intrinsic neural distinctions between word and object processing. To our knowledge, no study to date has been able to simultaneously control for both low-level visual form and higher-level components (e.g., phonology and semantics) while maintaining identical task demands in word versus object recognition; consequently, none has successfully identified the fundamental neural substrates that differentiate these processes.

Modern Chinese characters can be treated as a promising avenue for addressing these issues. This writing system preserve a selection of pictographs (Fig. 1, bottom of the two pictographs), each capable of dual recognition—as both a lexical item (e.g., a Chinese character "舟," meaning a boat) and a visual representation of the corresponding object (e.g., a line drawing of a boat). In the current study, we used artistically processed pictographs as stimuli for the functional magnetic resonance imaging (fMRI) experiments, ensuring that the stimuli were rigorously matched in terms of form, sound, and meaning between visual words and objects (Fig. 1; Fig. S1*a*). In addition to stimulus properties, we utilized a consistent set of common language tasks (realness judgment, sound retrieval, and meaning judgment) in the direct comparison between word and object recognition of pictographs to eliminate the influence of task demands (Fig. 2*a*). In each task, the same prompting paradigm was employed to guide participants to recognize pictographs as either words or objects. Specifically, pictographs were randomly interlaced and embedded within a continuous stream of real words (or real objects), thereby establishing contexts for recognizing these pictographs as words (or objects; Fig. 2*b*). With stimulus properties and task demands rigorously controlled, we ask: Do intrinsic neural differences persist between word and object recognition across the whole brain? If so, where are these differences located? Therefore, we focused primarily on the neural differences in activation intensities, activation patterns, and connectivity networks when pictographs were recognized as either words or objects.

Materials and Methods

Participants

This study recruited a total of 76 healthy university students. All participants were native Mandarin speakers, right-handed, and proficient in reading Chinese characters and had normal or corrected-to-normal vision. Among them, 36 participants [age, 25.4 (mean) ± 2.6 years (standard deviation, SD); 16 males] took part in the fMRI experiment. The remaining 40 subjects (age, 23.9 ± 2.4 years; 11 males) completed a material evaluation experiment. This experiment was used to assess and select materials for the fMRI study. This research was approved by the Ethics Committee of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University. All participants signed a written informed consent form prior to participating in the experiment.

Stimuli

Pictographs

In this study, we designed a novel set of visual stimuli that was meticulously crafted and rigorously evaluated. These materials possess several key characteristics:

1. The same visual stimulus can be flexibly recognized as either a word or an object without perceptual difficulty. When recognized as a word, the lines and curves of the visual stimulus are interpreted as the strokes or radicals of a Chinese character; when recognized as an object, these lines and curves are seen as the shape and appearance of an object (Fig. 1; Fig. S1*a*). This design enables a direct comparison between the neural mechanisms of visual word and object recognition, without the confounding effects of low-level visual feature differences (e.g., the pictograph shown in Fig. 1, the top one, when recognized as a word, represents the Chinese character "山" (mountain); when recognized as an object, it depicts a simple drawing of a mountain. The bottom pictograph, when recognized as a word, represents the Chinese character "舟" (boat); when recognized as an object, it represents a simple drawing of a boat).
2. The words and objects recognized convey the same concept. Specifically, the pronunciation of the word and the name of the object are identical; moreover, their meanings are also the same. This design eliminates the impact of sound and meaning differences in comparing visual recognition of words and objects. For example, in Figure 1, the top pictograph, when recognized as the Chinese character "山" (mountain), pronounced "/*shan1*/," represents a mountain. Similarly, when recognized as an image of a mountain, also named "/*shan1*/," it signifies a mountain as well. Likewise, the bottom pictograph, when recognized as the Chinese character "舟" (boat), pronounced "/*zhou1*/," represents the concept of a boat. When recognized as an image of a boat, also named "/*zhou1*/," it signifies a boat as well.
3. The words are real and widely used Chinese characters, while the objects are also real and common everyday items.

The word materials mentioned in this experiment are all currently commonly used Chinese characters. Chinese characters evolved from ancient pictographic characters, and many of these pictographic traits are still present in modern Chinese characters, especially in undecomposable Chinese characters. Undecomposable Chinese characters are those characters composed of strokes that cannot be further divided and can form compound characters. Most of them are pictographs evolved from drawings, with each character being a whole entity that represents independent meanings. The pictographs designed for this experiment are mostly undecomposable Chinese characters, with a few being compound characters (combinations of simple radicals) created artistically.

Fifty-five pictographs were designed and presented to 40 healthy university students for evaluation using seven-point Likert scale. The dimensions of evaluation were as follows:

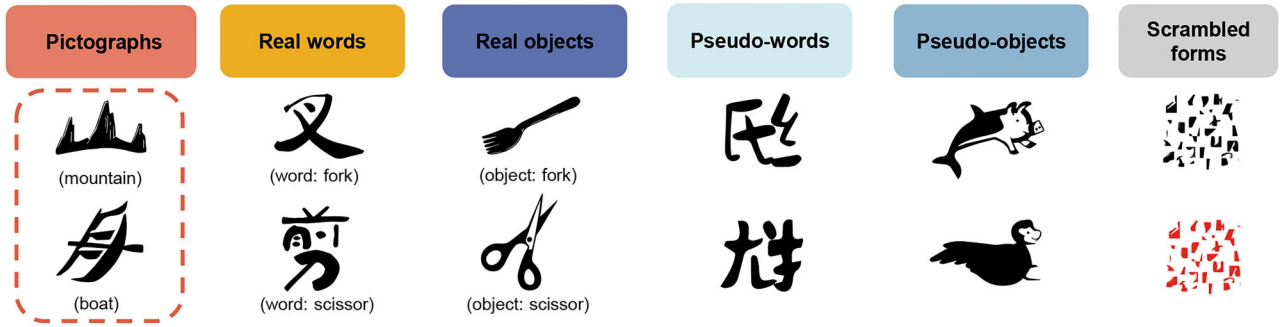

Figure 1. Sample stimuli for the fMRI experiment. Two examples of each type of stimulus are presented. This study mainly focuses on brain responses to pictographs, which can be recognized as both words and objects.

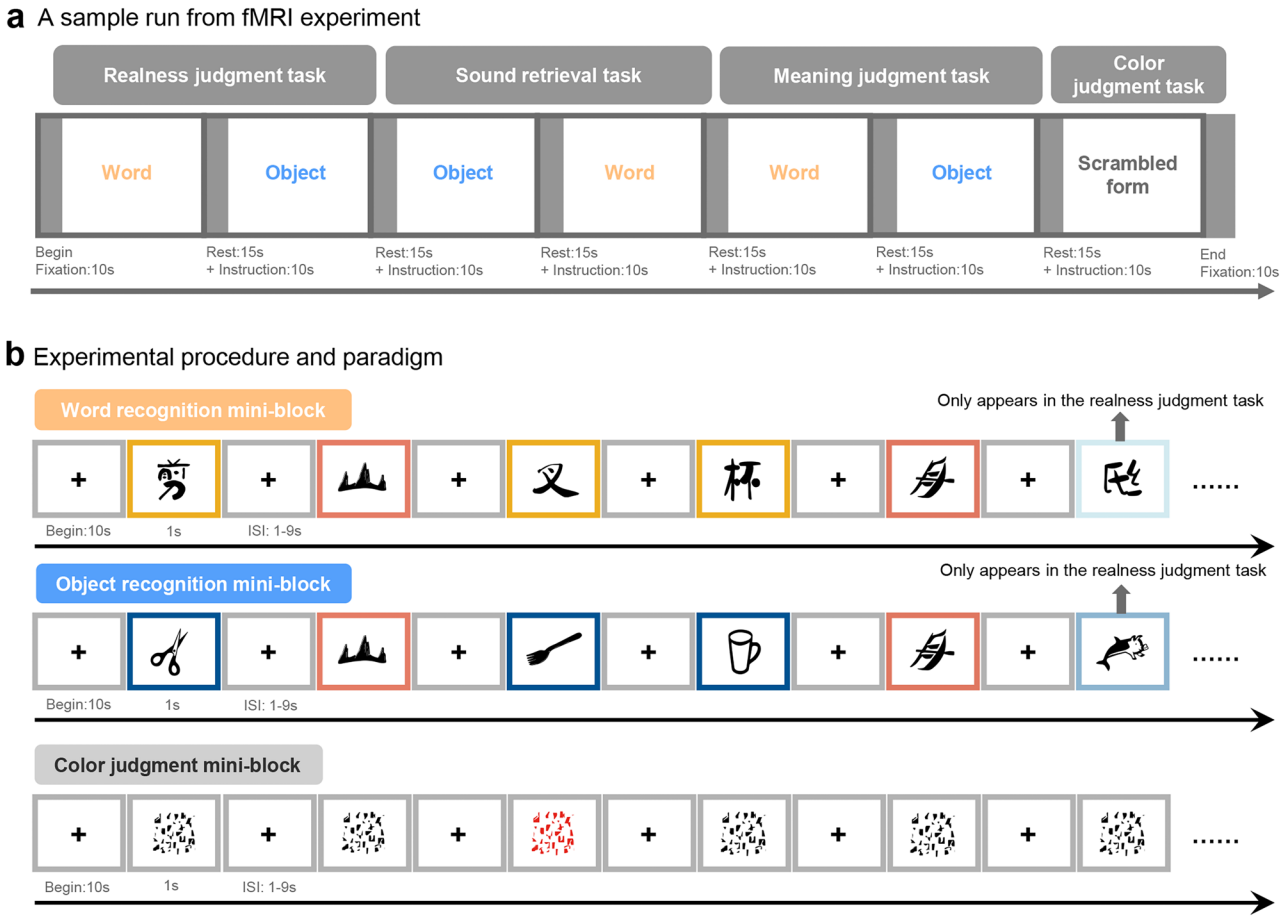

Figure 2. Experimental design. ***a***, A sample run from the fMRI experiment. Within each run, the participants sequentially completed three main tasks, followed by a baseline task (color judgment). Each task included a word and an object recognition miniblock, ensuring consistent task demands in comparing two recognition processes. The task and miniblock order were counterbalanced. ***b***, Examples of the experimental procedure and paradigm for miniblocks. The colors of the stimuli frames indicate the types of stimuli, corresponding to the colors shown in panel ***a***. These colored frames were added for illustration purposes only and were not present during the actual experiment. In the word (or object) recognition miniblock, pictographs were randomly interlaced and embedded within a continuous stream of real words (or objects), creating a word recognition (or object recognition) context, which guided participants to recognize pictographs as words (or objects). In these contexts, participants pressed buttons to judge whether each stimulus was a real word (or object) or a pseudoword (or pseudo-object; realness judgment task); silently read the word (or name the object; sound retrieval task); and judge whether the entity indicated by the word (or object) was animate (meaning judgment task). In the color judgment miniblock, participants pressed buttons to judge whether the color of each scrambled form was black or red. The detailed procedures are described in the Materials and Methods section.

1. The extent to which it can be recognized as a Chinese character (1, completely unrecognizable as a Chinese character, to 7, fully recognizable as a Chinese character) and type out the form of the character if recognizable
2. The extent to which it can be recognized as an object (1, completely unrecognizable as an object, to 7, fully recognizable as an object) and type out the name of the object if recognizable
3. The familiarity of the character/object in daily life (1, very unfamiliar, to 7, very familiar)

In the evaluation process, the order of evaluating the degree to which the stimuli could be recognized as Chinese characters or objects was balanced across participants. Half of the participants first evaluated the extent to which they could recognize the stimuli as characters, while

the other half first evaluated them as objects. Based on these evaluations, paired-sample *t* tests were conducted for each item. Ultimately, 20 visual stimuli were selected as the pictograph materials for fMRI experiment. There was no significant difference between the degree to which these stimuli were recognized as Chinese characters (mean, 6.229 ± SD, 0.590) and as objects [6.211 ± 0.623; $t_{(39)} = 0.321$; Cohen's *d* (hereafter abbreviated as "*d*") = 0.051; 95% confidence interval (hereafter abbreviated as "95% CI") = (−0.093, 0.128); $p = 0.750$; Bayes Factors (hereafter abbreviated as "$BF_{(10)}$") = 0.179; Fig. 3*a*]. The accuracy of recognizing them as characters was high (0.998 ± 0.011) and not significantly different from that of recognizing them as objects [0.992 ± 0.023; $t_{(39)} = 1.450$; $d = 0.229$; 95% CI (−0.002, 0.013); $p = 0.155$; $BF_{(10)} = 0.448$; Fig. 3*a*]. Additionally, the familiarity of the characters in daily life was high (6.821 ± 0.369) and not significantly different from the familiarity of the corresponding objects [6.733 ± 0.212; $t_{(39)} = 1.314$; $d = 0.208$; 95% CI (−0.048, 0.225); $p = 0.196$; $BF_{(10)} = 0.378$]. Thus, the selected pictograph materials showed balance in recognition as words and objects, with equal likelihood and extent of the same visual material being recognized as words and objects, and both words and objects being common and easily recognizable. The details of all 20 selected pictograph materials and the results of the paired-sample *t* tests for each material are presented in Text S1, Section A (Table S1). The selected materials were all concrete nouns, covering categories such as plants, animals, natural scenes, man-made objects, and body parts.

Real words

In the experiment, we designed 40 real word materials with visual properties (i.e., line thickness, texture, visual complexity, etc.) roughly matching those of the pictographs (Fig. 1; Fig. S1*b*). To ensure the real words visually resembled the pictographs as closely as possible, we utilized artistic fonts without affecting the recognizability of the real words. These materials could only be recognized as real words, not objects, and were used to create a "word recognition" environment in prompting paradigm. The real words are all concrete nouns, covering categories like plants, animals, and man-made objects. The university students participating in the material evaluation also assessed these 40 real word materials, following the same evaluation dimensions as previously mentioned. Eventually, we selected 30 real words as experimental materials. These 30 real word materials showed a significant difference in the extent of being recognized as Chinese characters (6.917 ± 0.197) versus as objects [1.401 ± 1.250; $t_{(39)} = 28.024$; $d = 4.431$; 95% CI (5.118, 5.915); $p < 0.001$; $BF_{(10)} = 1.442 \times 10^{24}$], with a high accuracy of recognition as characters (0.999 ± 0.005) and a high familiarity of the characters in daily life (6.759 ± 0.338; Fig. 3*a*). It is evident that these 30 real word materials could only be recognized as words, not as objects.

Real objects

Corresponding to the 30 real word materials, we designed 30 real object materials matched to their concepts while also ensuring a rough visual alignment with the pictographs (Fig. 1; Fig. S1*c*). To ensure the real objects visually resembled the pictographs as closely as possible, we artistically modified the line drawings of the real objects without affecting their recognizability. These materials could only be recognized as real objects, not as words, creating an "object recognition" environment in prompting paradigm. For these 30 real object materials, there was a significant difference in the extent of being recognized as Chinese characters (1.252 ± 0.871) and as objects [6.855 ± 0.198; $t_{(39)} = -40.457$; $d = -6.397$; 95% CI (−5.883, −5.323); $p < 0.001$; $BF_{(10)} = 1.019 \times 10^{30}$], with a high accuracy of recognition as objects (0.998 ± 0.012) and a high familiarity of the objects in daily life (6.531 ± 0.503; Fig. 3*a*). It is evident that these 30 real object materials could only be recognized as objects and not as words.

Pseudowords

In the experiment, we also created 10 pseudoword materials that roughly matched the visual properties of the pictographs (Fig. 1; Fig. S1*d*). These were created by deconstructing and rearranging components of real Chinese characters, resulting in forms resembling Chinese characters but not representing any real characters. These materials were only used for word miniblocks of realness judgment task. These 10 pseudowords were slightly recognizable as Chinese characters (3.680 ± 2.332) but not as objects [1.002 ± 0.016; $t_{(78)} = 7.262$; $d = 1.624$; 95% CI (1.943,

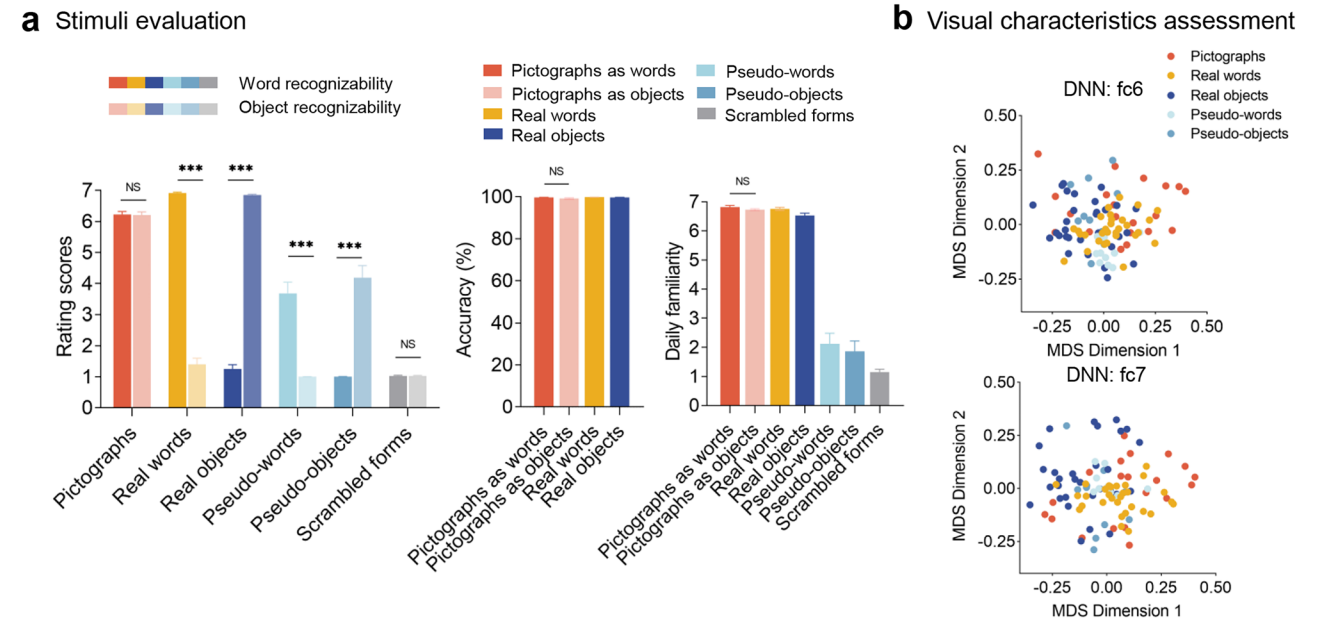

Figure 3. Stimuli assessment results. ***a***, Results of stimuli evaluation (sample size, $n = 40$). The left bar plot shows the extent to which each type of stimuli could be recognized as words (dark color) or objects (light color; 1, completely unrecognizable, to 7, fully recognizable). The middle bar plot shows the accuracy of recognition as words or objects (only pictographs recognized as words or objects, real words, and real objects have correct answers). The right bar plot shows the familiarity of each type of stimuli in daily life (1, very unfamiliar, to 7, very familiar). Results from the paired-sample *t* test revealed that pictographs were equally likely to be recognized as words or objects, with no significant difference in recognition accuracy or familiarity ratings between words and objects. Detailed results and descriptions for other types of stimuli are provided in the Materials and Methods section. The bars show the mean of each dimension of evaluation, and the error bars indicate the standard errors. NS, not significant; ***$p < 0.001$. ***b***, Results of the visual characteristic assessment for each type of stimulus using a deep neural network (DNN; i.e., Visual Geometry Group network, VGG-16). We extracted features from fully connected layers (i.e., fc6 and fc7) for each stimulus, used multidimensional scaling (MDS) for dimensionality reduction, and mapped them into a 2D space (each point represents a stimulus). The results show no significant group effect among the different types of stimuli in terms of global visual characteristics, ensuring the effective implementation of the experimental paradigm (i.e., embedding pictographs randomly among other stimulus types). See detailed results and descriptions in the Materials and Methods section and Text S1, Section A.

3.412); $p < 0.001$; $BF_{(10)} = 3.050 \times 10^7$], with low real-character familiarity (2.122 ± 2.159; Fig. 3*a*), indicating their resemblance to Chinese character forms without being real.

Pseudo-objects

Additionally, we designed 10 pseudo-objects with visual properties similar to the pictographs, created by deconstructing and arbitrarily reassembling parts of various animals, plants, and man-made objects (Fig. 1; Fig. S1*e*). These forms resembled objects but were not real existing objects. These materials were only used for object miniblocks of realness judgment task. These 10 pseudo-objects, slightly recognized as objects (4.188 ± 2.471) but not as Chinese characters [1.005 ± 0.032; $t_{(78)} = -8.145$; $d = -1.821$; 95% CI (−3.960, −2.405); $p < 0.001$; $BF_{(10)} = 1.245 \times 10^9$], with low real-object familiarity (1.868 ± 1.669; Fig. 3*a*), suggest a resemblance to objects without being real.

Scrambled forms

We also created 20 scrambled forms for the baseline task (color judgement task) by randomizing the pixels of pictographs (Fig. 1; Fig. S1*f*). These scrambled forms retained only low-level visual features and were devoid of meaning. Out of 20, 5 were randomly selected and colored red, while the rest were black. These 25 scrambled forms, serving as baseline materials, were used for the color judgment task. They were unrecognizable as either Chinese characters (1.030 ± 0.134) or objects (1.028 ± 0.104) and unfamiliar in daily life (1.152 ± 0.616; Fig. 3*a*).

All the aforementioned materials were manually drawn by researchers using Adobe Illustrator 2021 (https://www.adobe.com/products/illustrator.html) and Adobe Photoshop 2019 (https://www.adobe.com/products/photoshop.html). All materials were in black and white (450 × 450 pixels) and presented in a PNG format.

Experimental tasks and procedures

During fMRI scanning, the main experiment included three common language tasks and a baseline task. Each language task encompassed blocks for word and object recognition. The aim was to ensure consistency in task demands and difficulty when completing both word and object blocks, thereby minimizing the impact of task requirements on contrasting word and object recognition processes.

Since this study primarily focuses on the brain activity differences when pictographs are recognized as either words or objects, the overarching design principle for all language tasks is consistent. This involves creating implicitly guided contexts where participants are led to perceive the pictographs as either words or objects. Specifically, in the word (or object) recognition condition, pictographs are randomly interlaced and embedded within a continuous stream of real words (or objects), creating a word recognition (or object recognition) contexts, which guides participants to recognize pictographs as words (or objects; Fig. 2*b*).

In addition to the implicitly created environments, participants were also explicitly informed through instructions whether they were about to complete a word miniblock or an object miniblock. This combination of explicit instructions and implicit environmental guides enabled participants to flexibly recognize the same visual stimulus as either a word or an object.

Within such guided environments, participants completed three common language tasks during fMRI scanning: realness judgment, sound retrieval, and meaning judgment. Each task included both word and object miniblocks, enabling direct comparisons under matched task demands. Together, this task-based design allowed for informed and principled functional inference, as each task emphasized a distinct cognitive process relevant to language processing: visual form analysis, phonological access, and semantic retrieval, respectively. Additionally, a color judgment task was included as a baseline condition.

The order of the three language tasks was counterbalanced across runs to minimize order effects. The baseline color judgment task, however, was consistently placed at the end of each run. This design choice was intended to ensure that participants maintained full attention and cognitive engagement during the critical language tasks, which were the primary focus of the study.

Realness judgment task

This task emphasized visual form properties of the stimuli, differentiating real Chinese characters or objects from pseudowords or pseudo-objects. In this process, participants recognize pictographs as either real words or real objects. To ensure consistent task demands in word and object recognition, we used a word realness judgment task under the word recognition condition (word miniblock) and an object realness judgment task under the object recognition condition (object miniblock). This consistency in task demands is maintained during these cognitive processes. In the respective-guided environments, under word (object) recognition, 20 pictographs were randomly interlaced and embedded within a continuous stream of 30 real characters (objects), with an additional 10 pseudocharacters (pseudo-objects) randomly inserted. Participants used button presses to judge whether they were seeing real or pseudocharacters (objects). In this task, aside from the pseudocharacters and pseudo-objects, all other materials (pictographs, real characters, real objects) should be judged as real.

In the word context, this judgment may require fine-grained analysis of orthographic structure, such as stroke legality and radical positioning (Twomey et al., 2011; Carreiras et al., 2014; Liu et al., 2021). In contrast, object recognition may not involve such strict rule-based structural validation and may instead rely more on intuitive and holistic perception based on visual experience with object shapes (DiCarlo et al., 2012). Therefore, if a brain region shows stronger activation for words in this task, it may reflect increased demands on orthographic structural analysis supported by visuospatial attention. This involves the detailed analysis of the spatial structure of written words and the matching of these structures to internalized orthographic rules, such as standard stroke patterns and radical configurations, which are not required during object recognition.

In this and subsequent language tasks, the pseudorandom order and duration of stimuli in each of four runs were optimized using the optseq2 algorithm (https://surfer.nmr.mgh.harvard.edu/optseq/), with all participants following the same pseudorandom sequence. After optimization with the optseq2 algorithm, we manually selected the most suitable sequences from many optimized options to ensure effectiveness in the realness judgment task, with the following criteria: (1) the first two stimuli in each sequence that are always real; (2) no more than three consecutive pictographs; and (3) few occurrences (no more than three) of a pictograph following a pseudostimulus. These manually selected criteria aimed to ensure the effectiveness of pictographs within the implicit guiding context.

For these tasks, we used more real materials (30 items) than pictographs (20 items) to create a more robust guiding environment, thus using additional real materials as fillers. The number of pseudocharacters and pseudo-objects (10 items) was kept lower than that of pictographs because responses to pseudostimuli are not the experiment's primary focus; they merely serve as fillers to maintain engagement. This decision on the number of stimuli was made considering the cost-effectiveness of scanning duration.

Sound retrieval task

This task emphasized the sound properties of the stimuli, requiring participants to intentionally extract the sound from words or objects. In the word condition, they need to pronounce the words; in the object condition, they need to name the objects. To prevent head movement during vocal responses from affecting fMRI image quality, the task was adapted to silent reading and naming. Participants, without vocalizing, engaged in explicit cognitive processing to extract sound information and pressed a button with their right index finger at the onset of the silent reading or naming to record their response time. This task used only real items and pictographs, excluding pseudostimuli. In each miniblock, 20 pictographs were randomly interlaced and embedded within a continuous stream of 30 real characters (objects).

In this task, participants silently retrieved the phonological representation of each visual stimulus (e.g., mapping "舟" to /*zhou1*/) actively and maintained it in working memory, relying on efficient visual–phonological mapping. In the word condition, character–phonology associations are typically more automatic, direct, and rapid as a result of long-term experience with language use and reading, whereas object naming tends to be slower and more indirect, often relying on semantic mediation to access the corresponding phonological representation (Tan and Perfetti, 1999; Bell et al., 2001; Tsai et al., 2004; Oppermann et al.,

2010). Therefore, if a brain region shows stronger activation for words in this task, it may reflect a more efficient visual–phonological mapping process.

Meaning judgment task

This task emphasized meaning properties of the stimuli, requiring participants to intentionally extract the meaning of the experimental materials. In both word and object recognition conditions, participants judged the animate or inanimate nature of the materials via button presses. They assessed word meanings in the word miniblock and object meanings in the object miniblock (essentially evaluating the same concept). This task used only real items and pictographs, excluding pseudostimuli. In each miniblock, 20 pictographs were randomly interlaced and embedded within a continuous stream of 30 real characters (objects).

In this task, participants determined whether each stimulus referred to a living entity (e.g., whether "山" is animate), actively retrieving its meaning and maintaining it in working memory, relying on efficient visual-semantic mapping. Object recognition allows more direct and experience-based access to meaning through visual features such as shape, familiarity, or affordance (Arguin, 1996; Chiou and Lambon Ralph, 2016). In contrast, retrieving meaning from written characters typically involves more abstract and indirect form–meaning associations, shaped by language-based semantic knowledge (Nation, 2009; Fischer and Coello, 2015). This distinction may place greater demands on cognitive resources for accurate semantic retrieval and may require greater engagement of cognitive control mechanisms during word recognition compared with object recognition. Therefore, if a brain region shows stronger activation (or reduced deactivation) for words in this task, it may reflect increased involvement in top–down cognitive control over visual-semantic mapping.

Notably, although the baseline target rates differed between the realness and meaning tasks (5/6 vs 21/50), both frequentist and Bayesian analyses suggest that these differences did not meaningfully bias task performance or induce significant shifts in either discriminability or response bias (see Text S1, Section D for details; Green and Swets, 1966).

Color judgment task (baseline task)

This task served as the baseline task for the experiment. Baseline stimuli were scrambled forms created by pixelating pictographs, retaining only their low-level visual properties without any high-level linguistic information. In this task, participants simply judged the color of the scrambled forms. A few red scrambled forms (5 items) were randomly interspersed among black scrambled forms (20 items). Participants responded via button presses to indicate the color (black vs red).

Assessment of visual characteristics of stimuli

We utilized a commonly used visual convolutional neural network (i.e., Visual Geometry Group Network, VGG-16; Simonyan and Zisserman, 2015; Krizhevsky et al., 2017) to assess the visual characteristics of all five groups of novel stimuli we created (pictographs, real words, real objects, pseudowords, and pseudo-objects). We extracted the features of all stimuli from the fully connected layers of the convolutional neural network and used multidimensional scaling (MDS) analysis to reduce the dimensionality of these features and map them into a two-dimensional space. The results demonstrated that there was no significant grouping effect among the five groups of experimental materials in terms of high-level visual attributes, proving that the experimental paradigm's guiding contexts were effectively and seamlessly implemented under these mixed stimulus flow (Fig. 3*b*). Detailed descriptions of VGG-16 and the analysis results of the materials can be found in the Text S1, Section A.

Neuroimaging data acquisition

Participants were familiarized with the experimental materials and tasks the day before their MRI scans. They were able to accurately recognize the pictographs as either words or objects, and they had a clear understanding of the tasks to be performed in the fMRI experiment the following day.

The 36 participants in the fMRI main experiment at Beijing Normal University were scanned using a 3 T Siemens Magnetom Prisma scanner with a 64-channel phased–array head coil. Three types of images were acquired: 3D T1-weighted images, task-state fMRI images, and diffusion-weighted images (DWI; not used in this article). The task-state fMRI images included the main experiment of this study and a localizer experiment (data not used in this article). Considering the duration of the scan and potential fatigue from long scanning sessions, the scanning was conducted over 2 d, with an interval of no more than 3 d between scanning sessions. The main experiment comprised four runs, with two runs per scanning session, and these four runs were exact repetitions with only the order of tasks, blocks, and button presses balanced across different runs.

The first day included fMRI images for Run 1 and Run 2 of the main experiment, 3D T1-weighted images, and DWI scanning. The second day comprised fMRI images for Run 3 and Run 4 of the main experiment and two runs of the localizer experiment.

The procedure for a run in the main experiment was as follows (Fig. 2*a*,*b*, Run 1): realness judgment, sound retrieval, meaning judgment, and color judgment (baseline) tasks were sequentially conducted within a run. Each language task included both word and object blocks, with the next task appearing only after both blocks of the current task were completed.

The experiment used a miniblock event-related (ER) design, with all procedures presented using e-prime software (https://pstnet.com/products/e-prime/). Before each task, a 10 s instruction screen informed participants about the upcoming task and block, along with the corresponding button press [right index finger vs middle finger, e.g., in the word (object) block of the realness judgement task, real words (objects) corresponded to the index finger and pseudowords (pseudo-objects) to the middle finger or vice versa; the meaning judgment task also followed similar rules for button responses]. After the instructions, a 10 s fixation appeared, reminding participants to maintain attention as the task was about to start. Visual stimuli then appeared in the center of the screen for 1 s, during which participants could start their button-press response. After 1 s, the stimulus disappeared, followed by a jittered fixation (1–9 s). Each miniblock (word or object block) presented stimuli in a pseudorandom order and lasted ∼4.5 min on average. Button-press responses were allowed from stimulus appearance until the next one appeared. After each miniblock, there was a 15 s rest, with a screen prompt "Please rest for 15 s." After the rest, a 10 s instruction screen appeared for the next block. Following this process, participants completed the three language tasks. After all language tasks, they completed the baseline task, making color judgments (black vs red) on scrambled forms.

Each run lasted ∼30 min, with identical content across runs. The order of the three language tasks, the sequence of word and object miniblocks within each task, and the corresponding button presses were counterbalanced across different runs.

Structural 3D images in the sagittal plane were obtained with the following parameters: repetition time (TR), 2,530 ms; echo time (TE), 2.27 ms; flip angle (FA), 7°; field of view (FOV), 256×256 mm^2; slice number, 208; slice thickness, 1.0 mm; and voxel size, $1.0 \times 1.0 \times 1.0$ mm^3. The task fMRI data were collected in the transverse plane with a T2*-weighted echoplanar imaging sequence with the following parameters: TR, 2,000 ms; TE, 34 ms; FA, 90°; FOV, 200×200 mm^2; slice number, 72; slice thickness, 2 mm; and voxel size, $2.0 \times 2.0 \times 2.0$ mm^3. DWI was acquired in the transverse plane with the following parameters: TR, 3,900 ms; TE, 65 ms; FA, 90°; FOV, 256×256 mm^2; slice thickness, 2.0 mm, and voxel size, $2.0 \times 2.0 \times 2.0$ mm^3. There was a total of 64 diffusion weighting directions with a b value of 2,000 s/mm^2 and 10 b0 images. DWI was not further analyzed in the current study.

Postneuroimaging data acquisition

After completing all neuroimaging scans, each participant also completed the material stimulus evaluation form used in the assessment phase, to examine the 36 participants' recognition of the stimuli in the fMRI experiment (it should be noted that these 36 participants were not involved in the prior material assessment; a separate group of 40 individuals completed it). Among these 36 participants, there was no significant difference in the degree to which the 20 pictographs were recognized as Chinese characters (mean, $6.325 \pm$ SD, 0.462) versus as objects [6.283 ± 0.360; $t_{(35)} = 0.654$; $d = 0.109$; 95% CI (-0.088, 0.171); $p = 0.517$; $BF_{(10)} = 0.219$]. The accuracy of identifying them as characters was high (0.994 ± 0.016) and not

significantly different from that of identifying them as objects [0.997 ± 0.012; $t_{(35)} = -0.813$; $d = -0.135$; 95% CI (−0.010, 0.004); $p = 0.422$; $BF_{(10)} = 0.243$]. Moreover, the familiarity of the characters in daily life was high (6.626 ± 0.462) and not significantly different from the familiarity of the corresponding objects [6.560 ± 0.329; $t_{(35)} = 0.963$; $d = 0.161$; 95% CI (−0.074, 0.208); $p = 0.342$; $BF_{(10)} = 0.275$]. The details of all 20 selected pictograph materials and the results of the paired-sample t tests for each item are presented in Text S1, Section A (Table S2). These results indicate that for the participants in the experiment, the pictographs were recognized in a relatively balanced manner, aligning with our expectations. The postscan evaluation results, combined with a subsequent analysis based on brain activity pattern similarity (see details below, Experimental manipulation effectiveness analysis), indicated that the participants successfully recognized the pictographs as either words or objects as required during the scans, confirming the effectiveness of the experimental manipulation.

Neuroimaging data preprocessing

The fMRI data were analyzed with SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The first 10 s of each run were excluded from the analysis to allow for the initial stabilization of the fMRI signal. The preprocessing procedure included slice timing, motion correction, normalization to the Montreal Neurological Institute (MNI) space, and smoothing with an isotropic 6 mm full-width at half-maximum Gaussian kernel. For each subject, the data were first high-pass filtered and then fitted by a general linear model (GLM) with a boxcar regressor with a duration matching the response time to estimate the effect of the experimental conditions. Due to the prolonged duration of each run in the main experiment, we set a criterion that data with a maximum head rotation angle not exceeding 2 degrees and a translational displacement not exceeding 2 mm would be included in the final analysis. Based on these criteria, the data from one run of one participant (with a maximum translational displacement of 3.22 mm) was excluded from the subsequent analysis.

Whole-brain activation in word and object recognition

We combined the three language tasks for a group-level univariate activation analysis, aiming to observe overall brain activation when pictographs were recognized as words or objects above the baseline. In our GLM, we calculated the activation intensity (beta values) for each participant under each condition in 1-level analysis. Then, the activation intensity in each condition was averaged across all participants. Finally, the differences in the average intensity between word recognition and baseline (word recognition vs scrambled forms), as well as object recognition and baseline (object recognition vs scrambled forms), were compared using paired-sample t test. For multiple comparison corrections, we applied the False Discovery Rate method (FDR-corrected; $p < 0.05$; cluster size, >30 voxels). Importantly, unless specified otherwise, all analyses involving the word or object recognition condition in this paper refer to the recognition of pictographs as words or objects.

Differential activation effects of word and object recognition

Difference in activation intensity

For each voxel in the brain, we conducted a 3 (language tasks: realness judgement vs sound retrieval vs meaning judgement) × 2 (recognition type: word vs object recognition) within-subject repeated measures analysis of variance (ANOVA) to examine the overall activation differences in word and object recognition, with the dependent variable being the beta value in each voxel. In this analysis, we focused on the main effect of recognition type, aiming to identify brain regions reflecting the differences between word and object recognition. We used the Gaussian Random Field (GRF) method for multiple comparisons correction of the F-map (GRF corrected, voxel-level $p < 0.001$, cluster-level $p < 0.05$, one-tailed). For clusters that remained significant after correction, we saved them as masks to define regions of interest (ROIs) for this study.

We first extracted the average activation values within each ROI for both word and object recognition conditions (combining the three language tasks, after subtracting the baseline) and compared them (paired-sample t test, FDR-corrected $p < 0.05$) to observe the overall activation intensity differences within each ROI between the two conditions. Then, we conducted pairwise comparisons between word and object recognition (after subtracting the baseline) under each language task (paired-sample t test, FDR-corrected $p < 0.05$), identifying which language task mainly contributed to the overall difference effect. Before each paired-sample t test, we cleaned the data by removing outliers from the activation difference value (outside the interquartile range) to ensure the sample data conformed to a normal distribution.

Brain–behavior correlation

To further understand the functions and roles of these ROIs during task execution, we also extracted the average activation intensity within each ROI for each participant under each condition and performed Pearson's correlations with the behavioral performance (i.e., reaction times, RTs) of the participants for each corresponding task condition. Outlier detection using Cook's distance (Cook, 1977) revealed no data points with a Cook's distance greater than 1, indicating no significant outliers in the dataset.

Difference in activation pattern

For the ROIs defined in the aforementioned analyses, we conducted a multivariate level examination of the differences between word and object recognition. We employed multivariate pattern analysis (MVPA) to compare the similarity of activation patterns between word and object recognition. Initially, we acquired the whole-brain activation patterns (beta maps) for each trial. For each trial, a GLM was estimated for each voxel (using images normalized to MNI space but not smoothed), which included for each run one regressor containing the onsets of a given trial and another regressor containing the onsets of all other trials, both convolved with the canonical hemodynamic response function. This approach has been shown to be more representative of the true activation magnitudes unique to each trial type than other model estimation methods for rapid event-related designs (Mumford et al., 2012; Abdulrahman and Henson, 2016). Also included in the GLM were six head motion parameters and a global mean predictor of each run. The high-pass filter was set at 128 s. After model estimation, we obtained the whole-brain beta–weight image for each single trial.

To ensure reliable results, we used these beta maps for the following two sub-analyses in MVPA:

Leave-one-out support vector machine (SVM) analysis in ROIs. This analysis was used to investigate the neural pattern similarity in the ROIs between the word and object recognition conditions, focusing on the classification accuracy of activation patterns between conditions. This analysis was conducted with the Decoding Toolbox (Hebart et al., 2015) based on a library for SVM (Chang and Lin, 2011). It included the following steps.

(1) The beta values from unsmoothed beta maps of each voxel within the ROIs for each trial of word or object recognition condition were extracted. (2) The SVM classifier was trained to establish a predictive model using the beta values of a pair of conditions from three out of four runs (each condition containing all trials values for that condition). (3) The model's predictive ability (correct vs incorrect) was examined using the beta values of the same two conditions on the left-out run. This step was repeated four times, each time leaving out a different run (leave-one-out procedure). (4) The average classification accuracy was calculated across repetitions. Finally, the group-level decoding accuracy across subjects was assessed using a one-sample t test (FDR-corrected; $p < 0.05$; one-tailed).

Split-half correlation analysis in the ROIs. This analysis was used to examine the similarity of neural patterns in the ROIs between the two conditions (word vs object recognition), focusing on the stability of activation patterns across runs within a condition (Haxby et al., 2001). The data were split in half (i.e., half1 and half2). The hypothesis was that if the activation patterns between the two conditions were significantly different, the activation patterns produced by the two data halves from the same condition (i.e., "within-condition") would exhibit greater stability than the activation patterns produced by half1 from one condition and half2 from a different condition (i.e., "between-condition"). In other words, there was a higher correlation for the "within-condition" compared with the "between-condition" comparison.

This analysis included the following main steps. (1) For each subject, the preprocessed normalized fMRI data in the ROIs were divided into halves; this process was repeated to form all possible unique divisions. The four runs in each language task were split into two halves (each

containing two runs) repeated to form the two possible divisions (e.g., Way 1, Runs 1, 3 vs runs 2, 4; Way 2, Runs 1, 4 vs runs 2, 3). Given that Runs 1 and 2 and Runs 3 and 4 were scanned on separate days, to prevent potential adverse effects of scans conducted on different days, we excluded cases that each half of the data originated from the same day (e.g., Runs 1, 2 vs Runs 3, 4). Instead, each half of the data included data from both days. (2) The beta value of each voxel in each of the two conditions was extracted for each half of the data for a given division. The mean beta value across two conditions within the ROIs was subtracted from the beta value for one condition in each voxel to obtain a response more specific to that condition. As a result, there were four condition-specific beta values for each voxel and each subject. Each half of the data yielded two values corresponding to the two conditions (word vs object recognition). (3) For each pair of conditions, the condition-specific beta values of two conditions from one-half of the data were correlated with the condition-specific beta values of the other half across the voxels in the ROIs. This yielded four correlation coefficients: two within-condition coefficients (correlation between the two halves of the data from the same condition) and two between-condition coefficients (correlation between one-half of the data from one condition and the other half from a different condition). The four coefficients were further Fisher *Z*-transformed across subjects. (4) The difference value (*D* value) between the two conditions was calculated by subtracting the average value of two *Z*-transformed within-condition coefficients from the average value of two *Z*-transformed between–condition correlation coefficients. (5) The above steps were completed for all possible divisions, and the *D* values of all the divisions were averaged. Finally, the average *D* values were statistically tested (compared with zero) across subjects (one-sample *t* test; FDR-corrected; $p < 0.05$; one-tailed) to investigate whether significant differences emerged in the activation patterns between each pair of conditions. Considering data integrity, the participant with only three runs of data was excluded from this method's data analysis.

Experimental manipulation effectiveness analysis

In our fMRI experiment, we used explicit instructions and implicit environmental guides to manipulate participants to recognize the same pictographs as either words or objects in different conditions. To ensure the effectiveness of this experimental manipulation, beyond the behavioral evidence, we also tested it at the neural level. We hypothesized that perceiving pictographs as words would result in neural activation patterns similar to those of real words, and perceiving them as objects would result in activation patterns similar to those of real objects. We extracted and correlated the neural activation patterns for word and object recognition in each ROI, comparing these patterns to those of real words and real objects. The specific steps were as follows: (1) For each condition, a GLM was estimated for each voxel (using images normalized to MNI space but not smoothed) in each run. (2) Beta values for each condition were extracted from each ROI, with the beta values across voxels representing the activation pattern for that condition. (3) Within a specific language task, the activation pattern for pictographs word recognition in each run was Pearson-correlated with the activation pattern for real words and real objects, respectively. Similarly, the activation pattern for pictographs object recognition was correlated with real objects and real words, respectively. Correlation coefficients were Fisher *Z*-transformed and averaged across all runs. (4) For each language task, we obtained four Fisher *Z*-transformed correlation coefficients for each participant in each ROI: pictographs as words with real words, pictographs as words with real objects, pictographs as objects with real words, and pictographs as objects with real objects. (5) We conducted across-subject paired-sample *t* tests (FDR-corrected; $p < 0.05$; two-tailed) for the two correlation coefficients of pictograph word recognition and for those of pictograph object recognition, to examine whether the neural activation patterns in each ROI were similar to the manipulated direction of real stimuli.

Differences in the connectivity networks

In line with connectivity theory (Thomas Yeo et al., 2011; Park and Friston, 2013; Van Den Heuvel and Sporns, 2013), the function of a brain region often depends on its network. To further investigate the roles played by each key ROI in word and object recognition, we conducted an analysis of the differences in connectivity networks between core brain regions and other regions in the whole brain. We conducted seed-based whole-brain functional connectivity analysis. The specific process was as follows: (1) First, we obtained whole-brain beta maps for each trial using the GLM. The GLM setup was similar to MVPA, except that standardized and smoothed images were used here, facilitating the identification of relatively clustered clusters in whole-brain functional connectivity analysis. (2) We extracted the beta values of each voxel and averaged them by ROIs, for all trials in each condition using the BetA-Series COrrelation toolbox (BASCO v2.0; https://www.nitrc.org/projects/basco/; Göttlich et al., 2015). (3) The averaged beta value series of each ROI were correlated with the beta value series of each voxel in the whole brain across all trials for each condition. The Pearson's correlation coefficient reflected the functional connectivity strength between the ROI and the specific voxel in the condition (Rissman et al., 2004). (4) All correlation coefficients were Fisher *Z*-transformed. Each ROI served as a seed region (four seed ROIs defined in univariate activation analysis) for whole-brain functional connectivity analysis, resulting in a functional connectivity *Z*-map representing the connectivity between that ROI and all other voxels in the brain under a specific condition. (5) Using these *Z*-maps, for each voxel in the whole brain, we conducted a 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) $\times 2$ (recognition type, word vs object recognition) within-subject repeated–measure ANOVA to examine the overall differential connectivity effects of word and object recognition across the entire brain. The dependent variable was the connectivity strength (*Z* value) between that voxel and the seed ROI. (6) In this analysis, we focused on the main effect of recognition type, intending to identify connections across the entire brain that reflect the fundamental differences in word and object recognition. We used the GRF method for multiple comparison correction (GRF corrected; voxel-level $p < 0.001$; cluster-level $p < 0.05$; one-tailed). For clusters that remained significant after correction, we saved them as masks to serve as connectivity-regions of interest (Co-ROIs) in the connectivity analysis of this study. (7) For each Co-ROI, we extracted the connectivity values (*Z* values) within these regions of interest and conducted similar analyses as in the univariate activation analysis (paired-sample *t* test). The only difference is that the values used in the univariate activation analysis were activation strength (beta values), while here we used connectivity strength values (*Z* values) with the seed ROI. Based on the Co-ROIs identified in the connectivity analysis, their functions may indicate the primary role of each ROI.

Inferential statistics

We employed frequentist and Bayesian statistics for statistical tests. In addition to classical paired-sample *t* test, we also included Bayes factors (BFs) to assess the strength of evidence for each test. The BF represents the ratio of the probability of the data under the alternative hypothesis (indicating difference) to the probability of the data under the null hypothesis (Keysers et al., 2020). For instance, a BF of 10 signifies that the data are 10 times more likely under the alternative hypothesis than the null hypothesis. We computed frequentist statistics using the SPSS toolbox (Landau et al., 2004) and $BF_{(10)}$ using the JASP toolbox (Love et al., 2019).

Data availability

The datasets that support the findings of this study are available from the corresponding authors on request.

Results

Behavioral performance

Behavioral results revealed that in the realness judgment task, RTs for pictographs were slightly shorter in the word than the object recognition condition (Fig. 4*a*), marginally significant (word vs object recognition, 708 ± 105 vs 723 ± 118; $t_{(35)} = -2.07$; uncorrected $p = 0.046$; FDR-corrected $p = 0.10$); a similar marginally significant trend was observed in the sound retrieval task (word vs object recognition, 602 ± 147 vs 618 ± 133; $t_{(35)} = -1.75$; uncorrected $p = 0.09$; FDR-corrected $p = 0.10$). In the

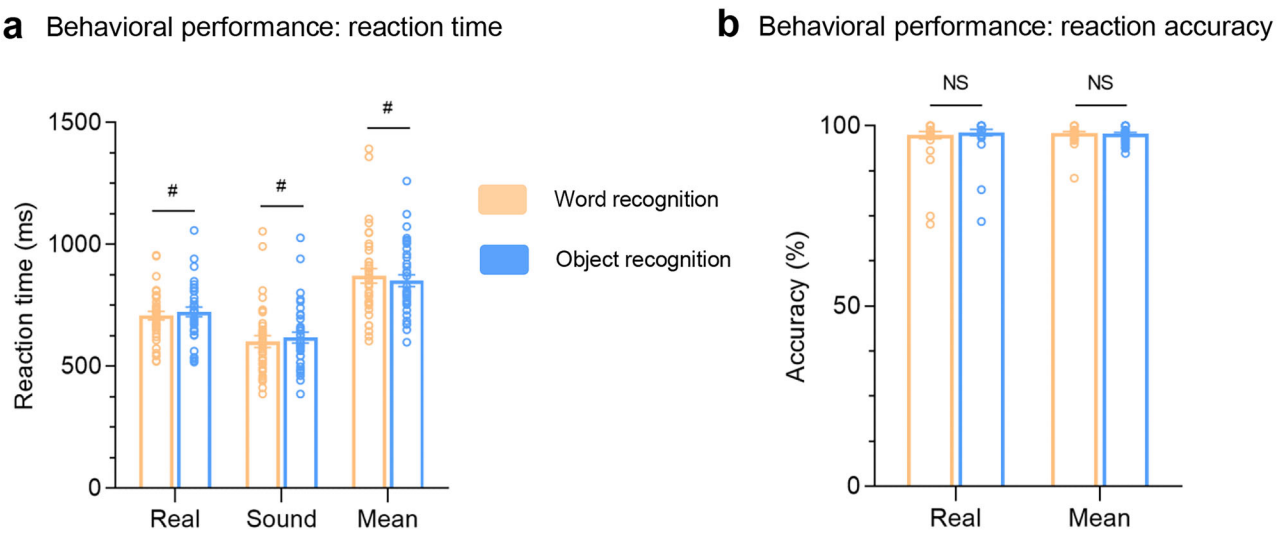

Figure 4. Behavioral performance in fMRI experiment. ***a***, Comparison of RTs (ms) between word and object recognition conditions for each main task (paired-sample *t* test; FDR-corrected; $^{\#}p \leq 0.1$). ***b***, Comparison of reaction accuracy (%) between word and object recognition conditions for each main task (paired-sample *t* test; FDR-corrected; NS, not significant). The bars represent the mean RT (panel ***a***) or accuracy (panel ***b***), with error bars indicating standard errors and each point representing individual participant data. Real, realness judgment task; Sound, sound retrieval task; Mean, meaning judgment task.

meaning judgment task, however, RTs were shorter in the object than the word recognition condition, though marginally significant (word vs object recognition, 871 ± 180 vs 851 ± 145; $t_{(35)} = 1.70$; FDR-corrected $p = 0.10$). These results are consistent with studies using real words or objects as materials in the same experimental tasks (Potter and Faulconer, 1975; Gerlach et al., 1999; Riesenhuber and Poggio, 2000; Rogers et al., 2004).

Regarding reaction accuracy, considering the experimental design, we could only collect accuracy rates for the realness judgment and meaning judgment tasks (sound retrieval involved silent reading and naming without vocalization, so the responses could not be recorded). The results showed that in both language tasks, accuracy rates were reaching the ceiling for both word and object recognition conditions, with no significant differences between them (Fig. 4*b*; word vs object recognition, realness judgment condition, 0.97 ± 0.06 vs 0.98 ± 0.05; $t_{(35)} = -1.79$; FDR-corrected $p = 0.17$; meaning judgment condition, 0.98 ± 0.03 vs 0.98 ± 0.02; $t_{(35)} = 0.22$; FDR-corrected $p = 0.83$). This indicates that the experimental tasks were relatively easy for the participants, who generally performed well.

Visual word and object recognition circuits in the whole brain

First, we conducted a whole-brain univariate activation analysis when pictographs were recognized as either words or objects, combining data from all three tasks to identify the visual word and object recognition circuits in the whole brain. The results revealed extensive activation across the brain during word recognition condition (Fig. 5*a*, left panel; Table 1) compared with the baseline. Additionally, we found that brain activation during object recognition (Fig. 5*a*, right panel; Table 1) largely mirrored that observed in the word recognition condition. Overlays of the two activation maps from Figure 5*a* revealed substantial overlap (Fig. 5*b*, red). A small number of voxels were activated during object recognition exclusively (Fig. 5*b*, purple), whereas almost no voxels were exclusively activated during word recognition (Fig. 5*b*, blue). The results suggest that when controlling for the identical forms, sounds, meanings, and tasks of words and objects, activated areas are broadly consistent between the two conditions.

Differences in activation intensities between word and object recognition

We conducted a 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) × 2 (recognition types, word vs object) within-subject repeated–measures ANOVA (analysis of variance) to examine the overall differences in whole-brain activation for pictographs. Since the contrast between word and object recognition is the core concern of this study, we focus primarily on the brain regions affected by the main effect of the recognition type and are not concerned with those affected by the main effect of language tasks. Additionally, brain regions from interactions are provided in Text S1, Section B, Figure S2.

The main effect of the recognition types yielded four almost symmetric clusters in both hemispheres (Fig. 5*c*, Table 2): the left IPL (lIPL), the right IPL (rIPL), the left ACC (lACC), and the right ACC (rACC).

For the bilateral IPL, word recognition showed significantly higher overall activation intensity than object recognition [lIPL, $t_{(35)} = 4.919$; $d = 0.820$, 95% CI (0.272, 0.653); $p < 0.001$; $BF_{(10)} = 1{,}059.103$; rIPL, $t_{(35)} = 5.064$; $d = 0.844$; 95% CI (0.196, 0.457); $p < 0.001$; $BF_{10} = 1{,}584.180$; Fig. 5*d*]. Specifically, for the lIPL, pairwise comparisons revealed that in both the realness [$t_{(32)} = 3.955$; $d = 0.688$; 95% CI (0.319, 0.995); $p = 0.001$; $BF_{(10)} = 73.405$] and sound tasks [$t_{(32)} = 2.826$; $d = 0.492$; 95% CI (0.125, 0.769); $p = 0.01$; $BF_{(10)} = 5.239$], higher activation intensity was detected for word recognition. However, the meaning task showed no significant difference [$t_{(32)} = 0.178$; $d = 0.031$; 95% CI (−0.303, 0.361); $p = 0.860$; $BF_{(10)} = 0.189$]. For the rIPL, higher activation intensity for word recognition was found in the realness task [$t_{(34)} = 5.202$; $d = 0.879$; 95% CI (0.360, 0.822); $p < 0.001$; $BF_{(10)} = 2{,}179.647$], which is consistent with the lIPL in this task. However, no significant difference was observed in meaning [$t_{(34)} = 1.091$; $d = 0.184$; 95% CI (−0.111, 0.369); $p = 0.424$; $BF_{(10)} = 0.314$] or sound task [$t_{(30)} = 0.643$; $d = 0.116$; 95% CI (−0.136, 0.262); $p = 0.525$; $BF_{(10)} = 0.232$].

For the bilateral ACC, word and object recognition showed negative activation compared with the baseline. The overall negative activation for word recognition was smaller in magnitude than that for object recognition (Fig. 5*d*; lACC, $t_{(35)} = 4.820$; $d =$

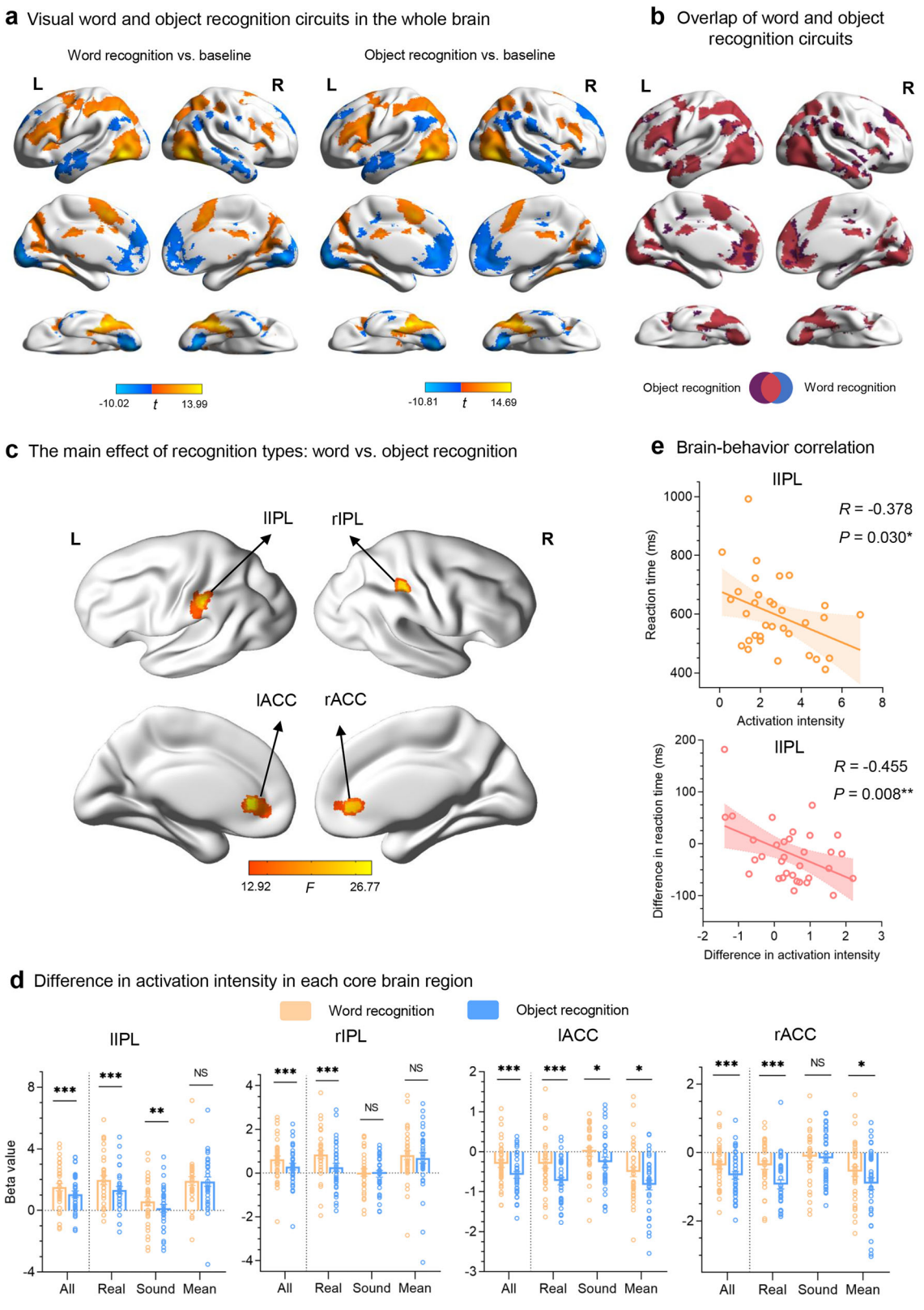

Figure 5. Whole-brain activation and differences in activation intensities between word and object recognition. ***a***, The whole-brain results for pictographs recognized as either words or objects above the baseline (combining three main tasks; paired-sample *t* test; FDR-corrected; $p < 0.05$; cluster size > 30 voxels; sample size, $n = 36$). For the full results of each cluster, see Table 1. ***b***, Spatial overlap of word and object recognition circuits. This panel illustrates the spatial relationship between brain regions activated during word recognition and object recognition, each compared with the baseline. It overlays the two activation maps from panel ***a*** (red, overlap of word and object recognition circuits; purple, object recognition circuit-specific voxels; blue, word recognition circuit-specific voxels). ***c***, Whole-brain differences in activation intensities between word and object recognition (sample size, $n = 36$). We conducted a 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) × 2 (recognition types, word vs object) within-subject repeated–measure ANOVA, and the main effect of recognition types yielded four almost symmetric clusters in both hemispheres: the lIPL (center coordinate, −54 −30 46; 82 voxels), the rIPL (center coordinate, 38 −36 46; 52 voxels), the lACC (center coordinate, −4 32 2; 81 voxels), and the rACC (center coordinate, 8 40 −2; 53 voxels; GRF corrected, voxel-level $p < 0.001$; cluster-level $p < 0.05$; one-tailed). ***d***, The ROI results for differences in activation intensities. The bars indicate the mean beta values for each

Table 1. Whole-brain activation results of the word and object recognition circuits (FDR corrected; $q < 0.05$; cluster size > 30 voxels)

Contrasts	Anatomical region of the peak voxel	Number of voxels	MNI coordinates of the peak voxel			Peak *t* value
			x	*y*	*z*	
Word recognition circuits:						
Word recognition > baseline	Right inferior temporal gyrus	25,145	48	−70	−6	13.987
	Right superior frontal gyrus	217	34	0	62	5.590
	Right insula	277	34	24	6	5.523
	Right Vermis7	682	2	−74	−24	5.331
	Left limbic lobe	428	−6	−30	30	5.269
	Right postcentral gyrus	327	42	−32	46	5.197
	Right inferior frontal gyrus	449	46	28	20	5.014
	Right inferior frontal operculum	233	40	6	34	4.159
	Right inferior frontal gyrus (orbital part)	32	28	32	−12	3.963
	Left hippocampus	35	−18	−34	0	3.903
Word recognition < baseline	Left lingual gyrus	4,149	−12	−86	−10	−10.022
	Left middle temporal pole	1,962	−54	14	−26	−5.999
	Right medial frontal gyrus (orbital part)	4,609	6	30	−10	−5.634
	Left supramarginal gyrus	696	−60	−56	26	−5.297
	Right middle temporal gyrus	1,640	56	−6	−24	−4.968
	Right angular gyrus	549	54	−56	26	−4.820
	Right middle cingulate gyrus	108	2	−24	44	−4.082
	Right medial pulvinar	32	10	−28	8	−3.931
	Right inferior frontal gyrus	32	54	42	−2	−3.838
	Right cerebellum crus I	186	24	−82	−28	−3.769
	Right superior medial frontal gyrus	32	8	28	66	−3.714
	Right supramarginal gyrus	102	68	−26	32	−3.705
	Left superior medial frontal gyrus	134	−8	28	64	−3.670
	Left post cingulate gyrus	60	−2	−48	32	−3.516
	Left cerebellum crus II	106	−20	−86	−32	−3.413
	Right precentral gyrus	44	50	−6	38	−3.210
Object recognition circuits:						
Object recognition > baseline	Left fusiform gyrus	8,961	−42	−52	−16	14.691
	Right inferior temporal gyrus	7,450	48	−70	−6	13.182
	Left supplementary motor area	6,821	−6	14	50	8.631
	Right inferior frontal gyrus (triangular part)	838	44	30	18	6.023
	Right cerebellum VIII	762	26	−72	−48	5.633
	Right insula	294	36	24	4	5.347
	Left anterior cingulate cortex	50	−8	22	30	4.821
	Left putamen	44	−24	−10	14	4.642
	Right superior frontal gyrus	60	34	0	62	4.640
	Right postcentral gyrus	177	46	−28	48	4.256
	Left pulvinar	49	−20	−32	0	4.244
	Left middle cingulate gyrus	45	−4	8	32	3.829
Object recognition < baseline	Right lingual gyrus	4,350	14	−78	−6	−10.813
	Right medial frontal gyrus (orbital part)	6,672	2	42	−6	−7.859
	Left superior temporal gyrus	2,332	−32	8	−14	−6.584
	Right superior temporal pole	2,246	42	24	−26	−6.127
	Right insula	442	40	−6	−8	−5.622
	Right middle cingulate gyrus	1,020	4	−26	44	−5.558
	Left angular gyrus	1,013	−60	−58	28	−5.095
	Right supramarginal gyrus1	798	62	−46	42	−4.786
	Right supramarginal gyrus2	448	60	−26	26	−4.755
	Left superior medial frontal gyrus	248	−10	32	56	−4.574
	Right cerebellum crus I	129	22	−80	−28	−3.965
	Left cerebellum crus II	333	−30	−82	−38	−3.947
	Right superior temporal gyrus	85	58	8	0	−3.760
	Left inferior temporal gyrus	37	−54	−28	−30	−3.697
	Left hippocampus	43	−22	−18	−16	−3.682
	Left middle temporal gyrus	55	−50	−34	−6	−3.475

←

ROI (after subtracting the baseline); the error bars indicate the standard errors, and each point indicates the data from a participant (paired-sample *t* test; FDR-corrected; NS, not significant; $*p \leq 0.05$; $**p \leq 0.01$; $***p \leq 0.001$). All, all three main tasks; Real, realness judgment task; Sound, sound retrieval task; Mean, meaning judgment task. ***e***, Brain–behavior correlation results. The top plot (yellow) shows a significant correlation between the RT and brain activation intensity in the lIPL in the sound task during the word recognition condition. The plot below (red) shows a significant correlation between RT differences and activation intensity differences (word vs object recognition) in the lIPL, as well as in the sound task. Outlier detection using Cook's distance (Cook, 1977) revealed no data points with Cook's distance greater than 1, indicating that significant outliers were not present in the dataset.

0.803; 95% CI (0.162, 0.397); $p < 0.001$; $BF_{(10)} = 805.972$; rACC, $t_{(35)} = 4.936$; $d = 0.823$; 95% CI (0.176, 0.421); $p < 0.001$; $BF_{(10)} = 1{,}110.327$). Specifically, word recognition produced a reduction in negative activation magnitude compared with object recognition in almost all tasks [lACC, realness, $t_{(35)} = 3.908$; $d = 0.651$; 95% CI (0.209, 0.661); $p = 0.001$; $BF_{(10)} = 70.329$; sound, $t_{(32)} = 2.324$; $d = 0.405$; 95% CI (0.032, 0.487); $p = 0.03$; $BF_{(10)} = 1.922$; meaning, $t_{(35)} = 2.555$; $d = 0.426$; 95% CI (0.068, 0.592); $p = 0.02$; $BF_{(10)} = 2.981$], except in the rACC during sound task, where no difference was found [rACC, realness, $t_{(33)} = 4.285$; $d = 0.735$; 95% CI (0.299, 0.840); $p < 0.001$; $BF_{(10)} = 176.467$; sound, $t_{(32)} = 0.364$; $d = 0.063$; 95% CI (−0.237, 0.340); $p = 0.718$; $BF_{(10)} = 0.198$; meaning, $t_{(35)} = 2.426$; $d = 0.404$; 95% CI (0.058, 0.656); $p = 0.03$; $BF_{(10)} = 2.305$].

We also conducted a correlation analysis between behavioral performance (i.e., RT) and brain activation intensity (Fig. 5*e*) within each of the four ROIs in each task condition to better understand the functions and roles of these ROIs during task execution. Significant results were observed only in the lIPL. A significant correlation was exhibited only in the sound task during the word recognition condition ($r = -0.378$; 95% CI (−0.639, −0.040); $p = 0.030$; $BF_{(10)} = 2.076$). Additionally, differences in RT between word and object recognition tasks were significantly correlated with differences in activation intensity ($r = -0.455$; 95% CI (−0.691, −0.133); $p = 0.008$; $BF_{(10)} = 6.473$; Fig. 5*e*).

In summary, even after controlling for low-level visual features, phonological and semantic properties of stimuli, as well as task demands, we still identified differences in the word and object recognition circuits within the bilateral IPL and ACC.

Table 2. Whole-brain activation intensity results for the main effect of recognition type (word vs object) from the ANOVA analysis: 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) × 2 (recognition types, word vs object; GRF corrected; voxel-level $p < 0.001$; cluster-level $p < 0.05$; one-tailed)

ANOVA	Anatomical region of the peak voxel	Number of voxels	MNI coordinates of the peak voxel			Peak *F* value
			x	*y*	*z*	
	lIPL	82	−54	−30	46	24.964
	rIPL	52	38	−36	46	23.131
	lACC	81	−4	32	2	26.774
	rACC	53	8	40	−2	22.894

Differences in activation patterns between word and object recognition

In the univariate activation analysis mentioned above, we identified four brain regions with differing activation intensities between the two processes. We now focused on differences in activation patterns within these regions. Here, we employed two complementary MVPA methods with different emphases (Fig. 6).

One method is SVM analysis, which focuses on the classification accuracy of activation patterns between conditions (Chang and Lin, 2011; Hebart et al., 2015; see Materials and Methods). We classified the activation patterns between word and object recognition across all three tasks combined and within each task separately for each ROI, finding significant differences in activation patterns between word and object recognition in the rIPL for the overall tasks [$t_{(35)} = 3.288$; $d = 0.548$; 95% CI (0.631, ∞); $p = 0.001$; $BF_{(10)} = 30.115$] and in the realness task [$t_{(35)} = 2.270$; $d = 0.378$; 95% CI (0.541, ∞); $p = 0.044$;

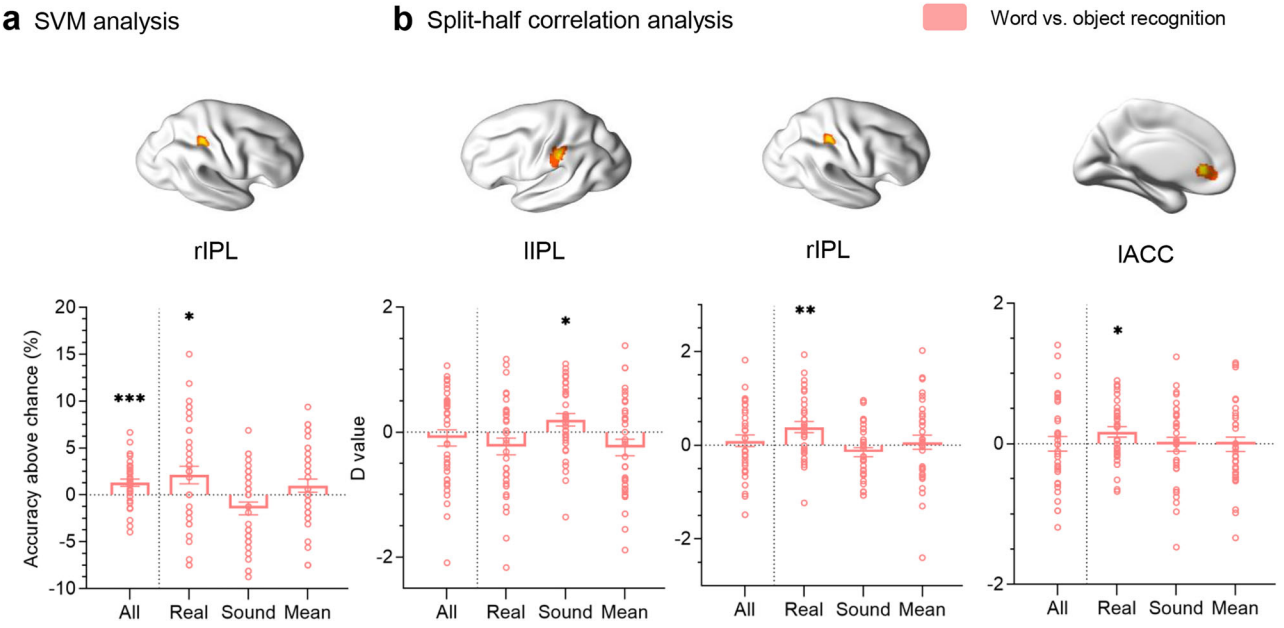

Figure 6. Differences in activation patterns between word and object recognition. ***a***, ROI results for differences in activation patterns using SVM analysis (for a detailed description of the calculation methods of the SVM, see the Materials and Methods section). The bars show the mean classification accuracy above chance (chance level: 50%) between the word and object recognition conditions in all three main tasks combined and each task separately. In this analysis, the activation patterns of the rIPL were significantly different between word and object recognition conditions in the combined tasks and the realness task. ***b***, ROI results for differences in activation patterns using the split-half correlation analysis. The bars show the mean difference value (*D* value) between the word and object recognition conditions in all three main tasks combined and each task separately (for a detailed description of the calculation methods for the *D* value, see the Materials and Methods section). In this analysis, the bilateral IPL and lACC showed significant differences in activation patterns between word and object recognition in the realness or sound tasks. The error bars indicate the standard errors, and each point indicate the data from a participant (one-sample *t* test; one-tailed; FDR-corrected; NS, not significant; $*p \leq 0.05$; $**p \leq 0.01$; $***p \leq 0.001$). The brain maps above the bar graphs show the locations of the ROIs defined in the main effect of the recognition types mentioned above. All, all three main tasks; Real, realness judgment task; Sound, sound retrieval task; Mean, meaning judgment task.

$BF_{(10)} = 3.363$], as reflected by above-chance classification accuracy for distinguishing word from object recognition (Fig. 6*a*).

Another method is split-half correlation analysis, which focuses on the stability of activation patterns across runs within a condition (Haxby et al., 2001; X. Luo et al., 2024; see Materials and Methods). We calculated the difference value (*D* value) of the activation patterns between word and object recognition. We observed a significant difference in the lIPL only in the sound task [Fig. 6*b*; $t_{(33)} = 1.969$; $d = 0.338$; 95% CI (0.028, ∞); $p_{uncorrected} = 0.029$; $BF_{10} = 1.977$]. For the rIPL, a significant effect was found in the realness task [$t_{(32)} = 3.291$; $d = 0.573$; 95% CI (0.188, ∞); $p = 0.004$; $BF_{(10)} = 29.451$], which was consistent with the findings from the SVM analysis and resembled the pattern observed in the univariate analysis. In the lACC, a significant effect was observed in the realness task [$t_{(32)} = 2.310$; $d = 0.402$; 95% CI (0.045, ∞); $p = 0.041$; $BF_{(10)} = 3.680$]. Besides this, no significant activation pattern differences were observed in the remaining ROIs or tasks ($ps > 0.05$).

In summary, even under strict control of stimulus properties and task demands, the IPL and ACC also exhibited differential neural activation patterns in certain tasks.

In addition, the results of experimental manipulation effectiveness analysis revealed that across the three tasks, in all ROIs, the similarity between the activation patterns of pictographs recognized as words and real words was significantly higher than the similarity between the activation patterns of pictographs recognized as words and real objects (pictographs as words with real words vs pictographs as words with real objects, correlation coefficients Fisher's *Z*, 0.29–0.83 vs 0.007–0.58; $ts_{(35)} = 3.96–18.14$; FDR-corrected; $ps < 0.001$). Similarly, the similarity between the activation patterns of pictographs recognized as objects and real objects was significantly higher than the similarity between the activation patterns of pictographs recognized as objects and real words (pictographs as objects with real objects vs pictographs as objects with real words, 0.31–0.82 vs 0.004–0.60; $ts_{(35)} = 5.35–16.59$; FDR-corrected; $ps < 0.001$; Fig. 7).

These results indicated that during the three tasks of the experiment, when the participants processed the pictographs as words or objects, the neural activation patterns were similar to those of the neural activation patterns of the real stimuli in the manipulated direction. This suggested that participants indeed perceived the pictographs as the specific category (words vs objects) as required by the experiment, demonstrating the effective manipulation of the experimental instructions.

Differences in connectivity networks between word and object recognition

To reveal broader neural networks associated with the core regions identified in our study and to examine whether their connectivity profiles differ between word and object recognition, we conducted whole-brain functional connectivity analyses using four key ROIs (bilateral IPL and ACC) as seed regions.

For each ROI, after obtaining connectivity maps for each condition, we performed a 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) × 2 (recognition types, word vs object) within-subject repeated–measure ANOVA on each voxel. Like the ANOVA in the univariate analysis, we focused on regions related to the main effect of recognition types, not language tasks, as differential networks between word and object recognition are our core concern. Connectivity analyses using regions identified from interactions in the univariate analysis as seeds are provided in Text S1 (Table S3, Fig. S5).

Seed region: lIPL

Using functional connectivity maps with the lIPL as the seed region, the main effect of the recognition types yielded eight significant clusters (Fig. 8*a*, left; Table 3): the left inferior temporal cortex (lIT), the right superior occipital gyrus (rSOG), the left superior frontal gyrus (lSFG), the right anterior temporal lobe (rATL), right medial prefrontal cortex (rmPFC), left medial prefrontal cortex (lmPFC), left mid-cingulate cortex (lMCC), and, notably, lACC, which is close to the left ACC region identified in the univariate analysis. No significant interaction effect was observed.

Post hoc tests for the main effect of the recognition types revealed that in all eight significant clusters, the connectivity strength was greater in the word recognition condition than in the object recognition condition (Fig. 8*a*, right panel). Statistical details are provided in Table 4.

Further detailed descriptions of the differences in the connectivity strengths across each task can be found in Text S1, Section C (Fig. S3, left panel). In general, the results obtained across all tasks showed a similar trend, even though the data from some brain regions were not sensitive enough to reach statistical significance in certain tasks.

Seed region: rIPL

For the rIPL, no significant clusters were identified for the main effect of the recognition types. However, a significant interaction effect was observed in the left superior occipital gyrus (lSOG; Fig. S3, right panel).

Seed region: lACC

For the lACC, the main effect of the recognition types yielded four significant clusters (Fig. 8*b*, left panel; Table 3). These regions are classic meaning processing regions (Billingsley, 2001; Bookheimer, 2002; Binder et al., 2009; Jensen et al., 2011; Wu et al., 2012): the left MTG (lMTG), left inferior frontal gyrus (triangular part; lIFG-tri), and left inferior frontal gyrus (opercular part; lIFG-oper). Notably, we also identified another lIPL, which is close to the left IPL region identified in the univariate analysis. No significant clusters were found for interaction effects.

Post hoc tests revealed that all four significant clusters had stronger connectivity with the seed region lACC in the word recognition task than in the object recognition task (Fig. 8*b*, right panel). Statistical details are provided in Table 4. Further differences in the connectivity strengths across each task can be found in Text S1, Section C (Fig. S4, left panel). In general, the results across all tasks showed a similar trend, even though the data from some brain regions were not sensitive enough to reach statistical significance in certain tasks.

Seed region: rACC

For the rACC, the main effect of recognition types yielded 10 significant clusters (Fig. 8*c*, left panel; Table 3; four of which were also found in the lACC connectivity results). These clusters included the left and right inferior temporal cortex (lIT and rIT) and right fusiform gyrus (rFG). Besides, the left middle frontal gyrus (lMFG, two clusters), left inferior frontal gyrus (triangular part; lIFG-tri, two clusters), left inferior frontal gyrus (opercular part; lIFG-oper), and lMTG are crucial regions for meaning processing (Billingsley, 2001; Bookheimer, 2002; Binder et al., 2009; Jensen et al., 2011; Wu et al., 2012). Notably, similar to the results in the lACC, another lIPL was found close to the left IPL region identified in the univariate analysis. No significant clusters were found for interaction effects.

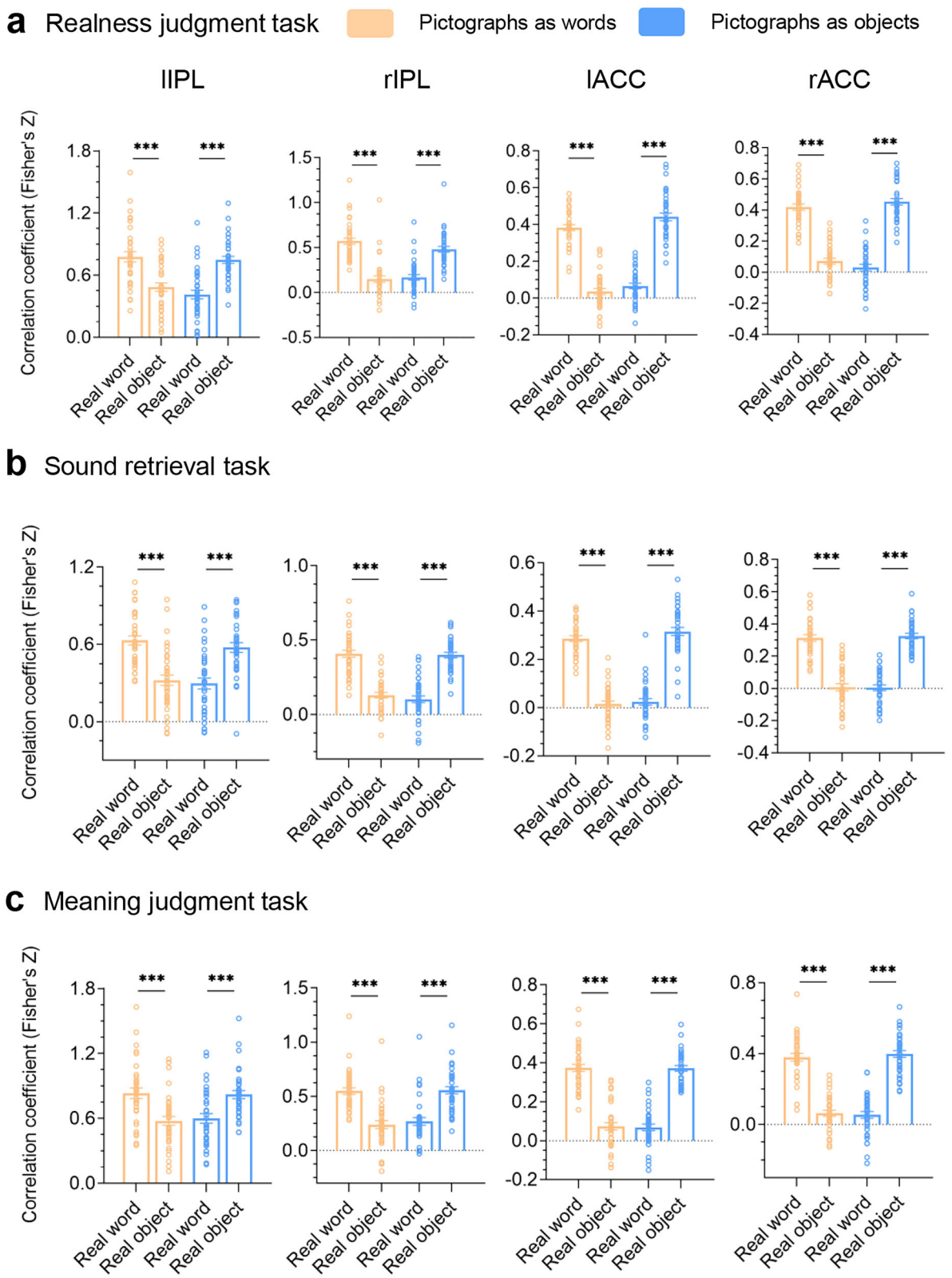

Figure 7. Experimental manipulation effectiveness in each task within each ROI. We hypothesize that when participants recognize pictographs as words, the neural activation patterns will be more similar to those of recognizing real words (compared with real objects) and vice versa for object recognition. The yellow bars indicate the similarity between neural activation patterns in each brain region when pictographs are recognized as words and the activation patterns when recognizing the stimuli type indicated by the label on the horizontal axis. The blue bars indicate the same similarity measure but for when pictographs are recognized as objects. Correlation coefficients were Fisher's *Z*-transformed, with error bars indicating standard errors and each point representing individual participant data. Results of the paired-sample tests indicate that subjects indeed recognize pictographs as the specific type (words vs objects) as manipulated (FDR-corrected; ***$p \leq 0.001$). ***a***, Results in the realness judgement task within each ROI. ***b***, Results in the sound retrieval task within each ROI. ***c***, Results in the meaning judgment task within each ROI.

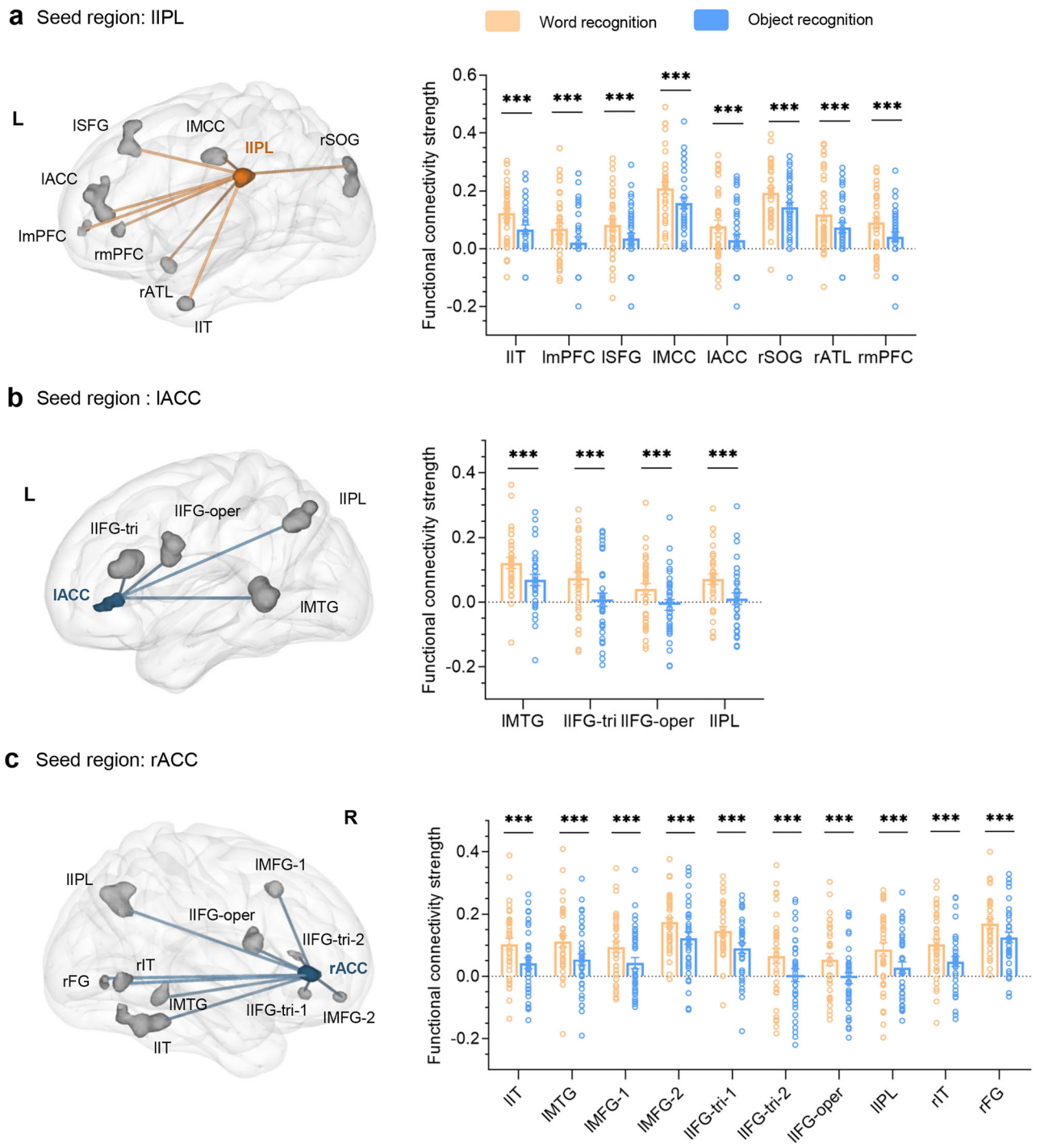

Figure 8. Differences in connectivity networks between word and object recognition. Using the core brain regions obtained from univariate analysis as seed regions (colored clusters in the left column brain maps), we performed seed-based whole–brain functional connectivity analysis for each condition. For the resulting whole-brain connectivity maps, we conducted a 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) × 2 (recognition types, word vs object) within-subject repeated–measures ANOVA. The main effect of recognition types yielded several brain regions distributed throughout the brain (gray clusters in the left column brain maps, referred to as Co-ROIs; GRF corrected, voxel-level $p < 0.001$; cluster-level $p < 0.05$; one-tailed). The connections between the seed region and Co-ROIs reflected differences in connectivity strength between word and object recognition (networks shown in the left column brain maps). The bar plots in the right column show the Co-ROI results of this analysis. The bars represent the mean functional connectivity strength between each Co-ROI and the corresponding seed region in each condition, with error bars indicating standard errors and each point representing individual participant data. These results demonstrate that the functional connectivity strengths between the Co-ROIs and the corresponding seed region are stronger in the word recognition condition than in the object recognition condition (paired-sample *t* test; FDR-corrected; $***p \leq 0.001$). ***a***, Results with lIPL as the seed region. ***b***, Results with lACC as the seed region. ***c***, Results with rACC as the seed region.

Post hoc tests showed that all 10 significant clusters had stronger connectivity with the seed region rACC in the word recognition task than in the object recognition task (Fig. 8*c*, right panel) Statistical details are provided in Table 4. Further differences in the connectivity strengths across each task can be found in Text S1, Section C (Fig. S4, right panel). In general, the results obtained across all tasks showed a similar trend, even though the data from some brain regions were not sensitive enough to reach statistical significance in certain tasks.

Table 3. Seed-based whole–brain functional connectivity results of the ANOVA analysis: 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) × 2 (recognition types, word vs object; GRF corrected; voxel-level $p < 0.001$; cluster-level $p < 0.05$; one-tailed)

ANOVA	Anatomical region of the peak voxel (Co-ROI)	Abbr.	Number of voxels	MNI coordinates of the peak voxel			Peak *F* value
				x	*y*	*z*	
Seed region: Left inferior parietal lobule		lIPL					
Main effect of recognition types (word vs object):	Left inferior temporal cortex	lIT	27	−42	0	−32	32.082
	Right superior occipital gyrus	rSOG	71	26	−90	28	26.278
	Left superior frontal gyrus	lSFG	62	−10	28	48	20.799
	Right anterior temporal lobe	rATL	23	52	12	−36	25.347
	Right medial prefrontal cortex	rmPFC	25	12	38	−4	18.969
	Left medial prefrontal cortex	lmPFC	20	0	58	−2	17.444
	Left middle cingulate cortex	lMCC	47	−2	−18	42	27.108
	Left anterior cingulate cortex	lACC	115	−2	46	8	27.220
Interaction effect of language tasks and recognition types:	None						
Seed region: Right Inferior parietal lobule		rIPL					
Main effect of recognition types (word vs object):	None						
Interaction effect of language tasks and recognition types:	Left superior occipital gyrus	lSOG	200	−10	−102	8	11.731
Seed region: Left anterior cingulate cortex		lACC					
Main effect of recognition types (word vs object):	Left middle temporal gyrus	lMTG	67	−60	−44	2	19.547
	Left inferior frontal gyrus (triangular part)	lIFG-tri	85	−48	26	22	21.526
	Left inferior frontal gyrus (opercular part)	lIFG-oper	36	−42	4	24	20.446
	Left inferior parietal lobule	lIPL	39	−32	−60	42	18.462
Interaction effect of language tasks and recognition types:	None						
Seed region: Right anterior cingulate cortex		rACC					
Main effect of recognition types (word vs object):	Left inferior temporal cortex	lIT	224	−56	−46	−12	22.006
	Right inferior temporal cortex	rIT	77	42	−56	−12	19.458
	Right fusiform gyrus	rFG	36	28	−68	−6	17.854
	Left middle frontal gyrus1	lMFG (cluster1)	64	−34	18	52	28.928
	Left middle frontal gyrus2	lMFG (cluster2)	30	−42	52	−4	18.770
	Left inferior frontal gyrus1 (triangular part)	lIFG-tri (cluster1)	33	−34	32	−2	21.080
	Left inferior frontal gyrus2 (triangular part)	lIFG-tri (cluster2)	30	−48	28	20	17.743
	Left inferior frontal gyrus (opercular part)	lIFG-oper	145	−40	4	24	34.368
	Left middle temporal gyrus	lMTG	67	−62	−46	2	17.005
	Left inferior parietal lobule	lIPL	216	−28	−64	44	20.255
Interaction effect of language tasks and recognition types:	None						

Abbr., abbreviation.

In summary, even under tightly controlled conditions, the strength of functional connectivity within IPL- and ACC-centered networks differs between word and object recognition, with both regions showing stronger connectivity with other brain areas specifically during word recognition.

Discussion

The central questions of this study are whether distinct neural mechanisms persist between visual word and object recognition after rigorously controlling for low-level visual features, phonological and semantic properties of stimuli, as well as task demands and, if so, which critical brain regions and functional networks underlie these differences. Using novel pictographic stimuli interpretable as both words and objects under identical tasks, we identified consistent neural differences involving the IPL, ACC, and their associated networks.

The potential functional role of IPL in visual recognition

Even under rigorously control, the brain still exhibits distinct neural mechanisms for visual word versus object recognition, with notably significant word-specific activation in the IPL. This specificity suggests a potential involvement of two core cognitive functions: orthographic visuospatial attention and automatic visual–phonological mapping.

In the word condition of the realness judgment task, participants were required to rapidly evaluate stimuli based on strict orthographic rules (e.g., stroke patterns, radical configurations). This process requires meticulous visuospatial attention to standardized structural features of word form and precise matching against a mental orthographic lexicon (Franceschini et al., 2012; Taylor et al., 2013; Carreiras et al., 2014). In contrast, object recognition relied more on general visual experience or three-dimensional interpretation than on two-dimensional standardized structural rules (Arguin, 1996; Chiou and Lambon Ralph, 2016). This fundamental difference likely accounts for IPL's significantly greater activation in word recognition conditions, reflecting specialized orthographic visuospatial attention. Converging evidence supports this interpretation. First, bilateral IPL exhibited greater activation during word recognition in realness task. Moreover, despite identical visual inputs, the rIPL showed robust multivariate differences between word and object recognition, suggesting that engagement by top–down orthographic visuospatial attention may functionally differentiate word from object recognition. Using the lIPL as seed, we observed stronger connectivity in word recognition with areas involved in high-level visual feature integration (lIT; Storrs et al., 2021; Graves et al., 2023; von Seth et al., 2023), visuospatial attention (rSOG; Zhou et al., 2021; Liu et al., 2022; Zheng et al., 2025), working memory buffering and goal maintenance

Table 4. Whole-brain functional connectivity results for the main effect of recognition type (word vs object) from the ANOVA analysis: 3 (language tasks, realness judgment vs sound retrieval vs meaning judgment) × 2 (recognition types, word vs object)

Seed region	Target region	*t*	*p*	95% CI	*d*	$BF_{(10)}$
lIPL	lIT	5.500	<0.001	[0.037, 0.080]	0.917	5,368.202
	rSOG	4.774	<0.001	[0.028, 0.069]	0.807	676.656
	rATL	4.702	<0.001	[0.027, 0.067]	0.784	583.497
	rmPFC	4.569	<0.001	[0.028, 0.072]	0.762	406.154
	lmPFC	4.599	<0.001	[0.025, 0.064]	0.767	440.484
	lSFG	5.314	<0.001	[0.030, 0.066]	0.886	3,188.068
	lMCC	5.902	<0.001	[0.036, 0.074]	1.012	13,772.05
	lACC	4.924	<0.001	[0.026, 0.063]	0.821	1,074.725
rIPL						
lACC	lMTG	5.752	<0.001	[0.034, 0.072]	1.033	6,804.699
	lIFG-tri	6.917	<0.001	[0.047, 0.086]	1.186	217,884.9
	lIFG-oper	4.613	<0.001	[0.027, 0.071]	0.769	456.884
	lIPL	5.966	<0.001	[0.040, 0.081]	1.023	16,396.62
rACC	lIT	5.624	<0.001	[0.039, 0.084]	0.951	7,011.724
	rIT	4.554	<0.001	[0.031, 0.080]	0.759	389.376
	rFG	4.488	<0.001	[0.024, 0.064]	0.748	326.138
	lMFG (cluster1)	4.849	<0.001	[0.030, 0.073]	0.808	873.978
	lMFG (cluster2)	4.08	<0.001	[0.026, 0.077]	0.68	109.953
	lIFG-tri (cluster1)	6.058	<0.001	[0.037, 0.073]	1.024	23,444.18
	lIFG-tri (cluster2)	7.043	<0.001	[0.044, 0.080]	1.208	305,491.3
	lIFG-oper	6.178	<0.001	[0.040, 0.079]	1.044	32,705.73
	lMTG	4.325	<0.001	[0.031, 0.085]	0.721	209.971
	lIPL	4.347	<0.001	[0.031, 0.087]	0.724	222.865

Degree of freedom was 35 for all contrasts. 95% CI, 95% confidence interval. $BF_{(10)}$ quantifies the evidence in favor of the alternative hypothesis compared with the null hypothesis; values >1 indicate increasing support for the alternative. lIPL, left inferior parietal lobule; rIPL, right inferior parietal lobule; lIT, left inferior temporal cortex; rIT, right inferior temporal cortex; rSOG, right superior occipital gyrus; rATL, right anterior temporal lobe; rmPFC, right medial prefrontal cortex; lmPFC, left medial prefrontal cortex; lSFG, left superior frontal gyrus; lMCC, left mid-cingulate cortex; lACC, left anterior cingulate cortex; rACC, right anterior cingulate cortex; rFG, right fusiform gyrus; lMFG, left middle frontal gyrus; lIFG-tri, left inferior frontal gyrus, triangular part; lIFG-oper, left inferior frontal gyrus, opercular part; lMTG, left middle temporal gyrus.

(bilateral mPFC and left AG; Baddeley, 2021; Lorenc et al., 2021; Nozari and Martin, 2024), and motivation selection/cognitive planning (MCC; Maldonado et al., 2020; Domic-Siede et al., 2021), indicating the lIPL may prioritize and integrate visuospatial structural information received from high-level visual areas via attentional mechanisms, transferring this information to the frontoparietal executive control networks (Menon and D'Esposito, 2022; Saylik et al., 2022). Collectively, the IPL, consistently implicated in attention networks (Markett et al., 2014; Numssen et al., 2021), may serve as a central node within a neural network specialized for orthographic visuospatial attention, which is critical for visual word processing.

The present findings raise an intriguing question: If the left IPL engages in orthographic visuospatial attention, why is its connectivity stronger with anterior temporal (e.g., IT) and dorsal visuospatial (e.g., rSOG) regions rather than with posterior ventral areas like the VWFA (Cohen et al., 2000; Lerma-Usabiaga et al., 2018)? One explanation is that our carefully matched pictograph stimuli eliminated low-level visual differences (e.g., line contours, curves), early visual areas (V1–V4), and posterior ventral stream regions likely processed both word and object conditions similarly. ROI analyses based on individually defined and canonical, anterior and posterior VWFA showed no differences between conditions (Text S1, Section E; Maldjian et al., 2003). This suggests the VWFA may support more basic and automated form encoding, whereas anterior IT and IPL may coordinate higher-order, top–down attention processes. Anterior ventral regions are known to integrate increasingly abstract features (Vinckier et al., 2007; Caffarra et al., 2021), and IPL's coupling with anterior IT may reflect structural matching against stored orthographic templates (Storrs et al., 2021; von Seth et al., 2023). Its dorsal connectivity with rSOG, a region involved in visuospatial attention (Zheng et al., 2025), further supports this interpretation. Notably, the lack of VWFA differences does not undermine VWFA's well-established role in word recognition (Dehaene and Cohen, 2011; Szwed et al., 2011; Bouhali et al., 2019; White et al., 2023). It is also possible that under rigorous visual controls, subtle orthographic distinctions between words and objects may escape current neuroimaging detection. Regardless, our findings highlight the interpretation that the neural specificity of word recognition may involve an IPL-driven network of orthographic visuospatial attention rather than being confined to the ventral visual cortex alone.

Beyond orthographic visuospatial attention, the left IPL also showed greater activation for words during the sound task, suggesting its involvement in visual–phonological mapping (Sliwinska et al., 2015; Bzdok et al., 2016; Y. Luo et al., 2024). In this task, participants silently read/named the visual symbol (e.g., viewing the character "舟" and internally producing */zhou1/*), requiring phonological generation and maintenance (Sliwinska et al., 2015; Bzdok et al., 2016; Y. Luo et al., 2024). Compared with objects, words allow more automatic visual–phonological mapping, engaging the IPL more strongly. Object naming, in contrast, relies on a slower, semantically mediated route (Arguin, 1996; Chiou and Lambon Ralph, 2016; Graves et al., 2023). This interpretation is further supported by a significant negative correlation between left IPL activation and response time in the word condition, indicating enhanced activity facilitates faster phonological retrieval. Distinct activation patterns in the lIPL between word and object conditions during the sound retrieval task, as revealed by multivariate analyses, may also suggest a preferential involvement of this region in the rapid mapping between visual forms and phonological representations, distinguishing it from object recognition. Functional connectivity showed enhanced coupling with regions involved in phonological working memory and executive control (mPFC, lSFG, and lMCC; Maldonado et al., 2020; Menon and D'Esposito, 2022; Saylik et al., 2022). Notably, the task involved silent reading/naming, requiring participants to internally generate and maintain phonological representations rather than overt articulation. Consequently, classical speech-related regions such as the superior temporal gyrus (STG) may not be significantly recruited here (Buchsbaum et al., 2001; Nozari and Martin, 2024; Zheng et al., 2025). Together, these findings support a potential role of the lIPL in efficient and automatic visual–phonological mapping, aligning with prior evidence (Graves et al., 2023; Nozari and Martin, 2024).

The potential functional role of ACC in visual recognition

The ACC is central to goal-directed cognitive control (Bush et al., 2000; MacDonald et al., 2000; Lambon Ralph et al., 2017; Maldonado et al., 2020). Its deactivation during tasks often reflects suppression of the default mode network (Greicius et al., 2003; Raichle, 2015; Sezer et al., 2022), while reduced deactivation indicates greater requirements for cognitive control. In our study, word recognition—particularly in realness and meaning judgment tasks—elicited less ACC deactivation than object recognition. In the realness task, participants were required to judge the authenticity of word stimuli based on strict two-dimensional structural rules, which may impose greater requirements on controlled attention and rule-based matching, thereby resulting in distinct activation strength and multivariate patterns

compared with object recognition. In the meaning judgment task, participants needed to extract and maintain semantic attributes, requiring active mapping between the visual form and conceptual meaning (Garrard et al., 2001; McRae et al., 2005). Notably, compared with objects whose visual features often afford more direct and experience-based access to meaning, written words rely on more abstract form–meaning associations (Noonan et al., 2013), typically requiring greater cognitive resources for accurate interpretation and integration. This distinction may underlie the increased reliance on ACC-mediated cognitive control during word recognition. Connectivity analyses further support this. During word recognition, the ACC exhibited stronger connectivity with the IFG, MTG, and other semantic regions (Whitney et al., 2011; Acheson and Hagoort, 2013; Carota et al., 2017; Zhao et al., 2021). These regions including the ACC are core components of the semantic control network and have been consistently implicated in the controlled retrieval and integration of semantic information (Bush et al., 2000; MacDonald et al., 2000; Lambon Ralph et al., 2017). Converging evidence from activation strength, multivariate pattern differences, and connectivity analyses collectively underscores the potential role of the ACC in top–down cognitive control over the mapping between visual form and semantic representations.

Collaborative networks of IPL and ACC

While IPL and ACC support distinct functions, connectivity analyses reveal their collaboration. Specifically, seed-based analyses revealed significant coupling between the IPL and ACC: using the IPL as the seed showed increased connectivity with the ACC and vice versa. This collaboration underscores the coordinated processing and integration of multiple types of language information (form, phonology, semantics) throughout word and object recognition. Although we have emphasized the distinct potential functions of the IPL and ACC based on task, this distinction does not imply that these regions and their networks operate in isolation. For instance, the IPL network also showed increased connectivity with the right ATL, a key region for semantic processing, while the ACC exhibited significant univariate activation during the sound retrieval task, possibly implicating involvement in phonological processing. These findings indicate that, rather than functioning as isolated modules, the IPL and ACC form an integrated network that facilitates top–down and bottom–up interactions, enabling flexible and efficient processing of visual symbols.

Limitations

First, while our use of simple, single-character pictographs allowed for rigorous control over visual, phonological, and semantic features, this design inherently excluded combinatorial linguistic processes fundamental to natural reading, such as letter sequencing, morphological decomposition, and multicharacter integration (Lothaire, 1997; Berstel and Perrin, 2007). The absence of these language-specific operations, which engage distinct neurocognitive mechanisms (e.g., grapheme clustering, syntactic parsing; Grodzinsky et al., 2021; Pattamadilok and Sato, 2022), limits the direct comparability of our findings with studies involving alphabetic scripts or more complex logographic characters. Future research using multiletter words or compound characters is needed to determine how the basic mechanisms identified here generalize to more ecologically valid reading conditions. Second, while not a major concern, we note that a minimal degree of residual hemodynamic influence from adjacent trials may still be present despite the use of jittered timing and trial-wise modeling. Future studies may consider incorporating longer interstimulus intervals or model-based deconvolution techniques to further reduce potential overlap.

Conclusion

By rigorously matching stimulus properties (i.e., visual form, phonology and semantics) and task demands between visual word and object recognition using Chinese pictographs, this study demonstrated that the IPL, ACC, and their functional networks underlie the core neural basis of the distinction between word and object recognition processing.

References

Abdulrahman H, Henson RN (2016) Effect of trial-to-trial variability on optimal event-related fMRI design: implications for beta-series correlation and multi-voxel pattern analysis. Neuroimage 125:756–766.
Acheson DJ, Hagoort P (2013) Stimulating the brain's language network: syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. J Cogn Neurosci 25:1664–1677.
Arguin M (1996) Shape integration for visual object recognition and its implication in category-specific visual agnosia. Vis Cogn 3:221–276.
Baddeley AD (2021) Developing the concept of working memory: the role of neuropsychology. Arch Clin Neuropsychol 36:861–873.
Bell BD, Hermann BP, Woodard AR, Jones JE, Rutecki PA, Sheth R, Dow CC, Seidenberg M (2001) Object naming and semantic knowledge in temporal lobe epilepsy. Neuropsychology 15:434–443.
Berstel J, Perrin D (2007) The origins of combinatorics on words. Eur J Comb 28:996–1022.
Billingsley RL (2001) Functional MRI of phonological and semantic processing in temporal lobe epilepsy. Brain 124:1218–1227.
Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex 19:2767–2796.
Bookheimer S (2002) Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu Rev Neurosci 25:151–188.
Bouhali F, Bézagu Z, Dehaene S, Cohen L (2019) A mesial-to-lateral dissociation for orthographic processing in the visual cortex. Proc Natl Acad Sci U S A 116:21936–21946.
Bright P, Moss H, Tyler LK (2004) Unitary vs multiple semantics: PET studies of word and picture processing. Brain Lang 89:417–432.
Buchsbaum BR, Hickok G, Humphries C (2001) Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cogn Sci 25:663–678.
Bush G, Luu P, Posner MI, Bush G, Luu P, Posner MI, Bush G, Luu P, Posner MI (2000) Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci 4:215–222.
Bzdok D, Hartwigsen G, Reid A, Laird AR, Fox PT, Eickhoff SB (2016) Left inferior parietal lobe engagement in social cognition and language. Neurosci Biobehav Rev 68:319–334.
Caffarra S, Martin CD, Lizarazu M, Lallier M, Zarraga A, Molinaro N, Carreiras M (2017) Word and object recognition during reading acquisition: MEG evidence. Dev Cogn Neurosci 24:21–32.
Caffarra S, Karipidis II, Yablonski M, Yeatman JD (2021) Anatomy and physiology of word-selective visual cortex: from visual features to lexical processing. Brain Struct Funct 226:3051–3065.
Carota F, Kriegeskorte N, Nili H, Pulvermüller F (2017) Representational similarity mapping of distributional semantics in left inferior frontal, middle temporal, and motor cortex. Cereb Cortex 27:294–309.
Carreiras M, Armstrong BC, Perea M, Frost R (2014) The what, when, where, and how of visual word recognition. Trends Cogn Sci 18:90–98.
Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:1–27.
Chiou R, Lambon Ralph MA (2016) The anterior temporal cortex is a primary semantic source of top-down influences on object recognition. Cortex 79:75–86.
Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123:291–307.

Cohen L, Dehaene S (2004) Specialization within the ventral stream: the case for the visual word form area. Neuroimage 22:466–476.
Cook RD (1977) Detection of influential observation in linear regression. Technometrics 19:15–18.
Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. Trends Cogn Sci 15:254–262.
Devereux BJ, Clarke A, Marouchos A, Tyler LK (2013) Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. J Neurosci 33:18906–18916.
DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434.
Domic-Siede M, Irani M, Valdés J, Perrone-Bertolotti M, Ossandón T (2021) Theta activity from frontopolar cortex, mid-cingulate cortex and anterior cingulate cortex shows different roles in cognitive planning performance. Neuroimage 226:117557.
Fischer M, Coello Y (2015) *Conceptual and interactive embodiment: foundations of embodied cognition*. London: Routledge.
Franceschini S, Gori S, Ruffino M, Pedrolli K, Facoetti A (2012) A causal link between visual spatial attention and reading acquisition. Curr Biol 22: 814–819.
Garrard P, Lambon Ralph MA, Hodges JR, Patterson K (2001) Prototypicality, distinctiveness, and intercorrelation: analyses of the semantic attributes of living and nonliving concepts. Cogn Neuropsychol 18:125–174.
Gerlach C, Law I, Gade A, Paulson OB (1999) Perceptual differentiation and category effects in normal object recognition. Brain 122:2159–2170.
Giari G, Leonardelli E, Tao Y, Machado M, Fairhall SL (2020) Spatiotemporal properties of the neural representation of conceptual content for words and pictures – an MEG study. Neuroimage 219:116913.
Göttlich M, Beyer F, Krämer U (2015) BASCO: a toolbox for task-related functional connectivity. Front Syst Neurosci 9:126.
Graves WW, Purcell J, Rothlein D, Bolger DJ, Rosenberg-Lee M, Staples R (2023) Correspondence between cognitive and neural representations for phonology, orthography, and semantics in supramarginal compared to angular gyrus. Brain Struct Funct 228:255–271.
Green DM, Swets JA (1966) *Signal detection theory and psychophysics*, pp xi, 455. New York: John Wiley.
Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci U S A 100:253–258.
Grodzinsky Y, Pieperhoff P, Thompson C (2021) Stable brain loci for the processing of complex syntax: a review of the current neuroimaging evidence. Cortex 142:252–271.
Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.
Hebart MN, Görgen K, Haynes J-D (2015) The decoding toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front Neuroinform 8:88.
Jensen EJ, Hargreaves IS, Pexman PM, Bass A, Goodyear BG, Federico P (2011) Abnormalities of lexical and semantic processing in left temporal lobe epilepsy: an fMRI study. Epilepsia 52:2013–2021.
Kay KN, Yeatman JD (2017) Bottom-up and top-down computations in word- and face-selective cortex. Elife 6:e22341.
Keysers C, Gazzola V, Wagenmakers E-J (2020) Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nat Neurosci 23:788–799.
Klaus J, Hartwigsen G (2019) Dissociating semantic and phonological contributions of the left inferior frontal gyrus to language production. Hum Brain Mapp 40:3279–3287.
Köhler S, Moscovitch M, Winocur G, McIntosh AR (2000) Episodic encoding and recognition of pictures and words: role of the human medial temporal lobes. Acta Psychol 105:159–179.
Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Communications of the ACM 60: 84–90.
Kuhnke P, Chapman CA, Cheung VKM, Turker S, Graessner A, Martin S, Williams KA, Hartwigsen G (2023) The role of the angular gyrus in semantic cognition: a synthesis of five functional neuroimaging studies. Brain Struct Funct 228:273–291.
Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. Nat Rev Neurosci 18:42–55.
Landau S, Everitt B, Everitt BS (2004) *A handbook of statistical analyses using SPSS*. Boca Raton, FL: Chapman & Hall/CRC.
Lerma-Usabiaga G, Carreiras M, Paz-Alonso PM (2018) Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. Proc Natl Acad Sci U S A 115:E9981–E9990.
Li M, Xu Y, Luo X, Zeng J, Han Z (2020) Linguistic experience acquisition for novel stimuli selectively activates the neural network of the visual word form area. Neuroimage 215:116838.
Liu T, Thiebaut de Schotten M, Altarelli I, Ramus F, Zhao J (2022) Neural dissociation of visual attention span and phonological deficits in developmental dyslexia: a hub-based white matter network analysis. Hum Brain Mapp 43:5210–5219.
Liu Y, Shi G, Li M, Xing H, Song Y, Xiao L, Guan Y, Han Z (2021) Early top-down modulation in visual word form processing: evidence from an intracranial SEEG study. J Neurosci 41:6102–6115.
Lorenc ES, Mallett R, Lewis-Peacock JA (2021) Distraction in visual working memory: resistance is not futile. Trends Cogn Sci 25:228–239.
Lothaire M (1997) *Combinatorics on words*. Cambridge, UK: Cambridge University Press.
Love J, et al. (2019) JASP: graphical statistical software for common statistical designs. J Stat Softw 88:1–17.
Luo X, Li M, Zeng J, Dai Z, Cui Z, Zhu M, Tian M, Wu J, Han Z (2024) Mechanisms underlying category learning in the human ventral occipitotemporal cortex. Neuroimage 287:120520.
Luo Y, Wang K, Jiao S, Zeng J, Han Z (2024) Distinct parallel activation and interaction between dorsal and ventral pathways during phonological and semantic processing: a cTBS-fMRI study. Hum Brain Mapp 45:e26569.
MacDonald AW, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288:1835–1838.
Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19:1233–1239.
Maldonado IL, Parente de Matos V, Castro Cuesta TA, Herbet G, Destrieux C (2020) The human cingulum: from the limbic tract to the connectionist paradigm. Neuropsychologia 144:107487.
Markett S, Reuter M, Montag C, Voigt G, Lachmann B, Rudorf S, Elger CE, Weber B (2014) Assessing the function of the fronto-parietal attention network: insights from resting-state fMRI and the attentional network test. Hum Brain Mapp 35:1700–1709.
McRae K, Cree GS, Seidenberg MS, Mcnorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. Behav Res Methods 37:547–559.
Menard MT, Kosslyn SM, Thompson WL, Alpert NM, Rauch SL (1996) Encoding words and pictures: a positron emission tomography study. Neuropsychologia 34:185–194.
Menon V, D'Esposito M (2022) The role of PFC networks in cognitive control and executive function. Neuropsychopharmacology 47:90–103.
Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643.
Nation K (2009) Form–meaning links in the development of visual word recognition. Philos Trans R Soc B Biol Sci 364:3665–3674.
Noonan KA, Jefferies E, Visser M, Lambon Ralph MA (2013) Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. J Cogn Neurosci 25:1824–1850.
Nozari N, Martin RC (2024) Is working memory domain-general or domain-specific? Trends Cogn Sci 28:1023–1036.
Numssen O, Bzdok D, Hartwigsen G (2021) Functional specialization within the inferior parietal lobes across cognitive domains. Elife 10:e63591.
Oppermann F, Jescheniak JD, Schriefers H, Görges F (2010) Semantic relatedness among objects promotes the activation of multiple phonological codes during object naming. Q J Exp Psychol 63:356–370.
Park H-J, Friston K (2013) Structural and functional brain networks: from connections to cognition. Science 342:1238411.
Pattamadilok C, Chanoine V, Pallier C, Anton J-L, Nazarian B, Belin P, Ziegler JC (2017) Automaticity of phonological and semantic processing during visual word recognition. Neuroimage 149:244–255.
Pattamadilok C, Sato M (2022) How are visemes and graphemes integrated with speech sounds during spoken word recognition? ERP evidence for supra-additive responses during audiovisual compared to auditory speech processing. Brain Lang 225:105058.
Potter MC, Faulconer BA (1975) Time to understand pictures and words. Nature 253:437–438.

Price CJ, McCrory E, Noppeney U, Mechelli A, Moore CJ, Biggio N, Devlin JT (2006) How reading differs from object naming at the neuronal level. Neuroimage 29:643–648.
Raichle ME (2015) The brain's default mode network. Annu Rev Neurosci 38: 433–447.
Riesenhuber M, Poggio T (2000) Models of object recognition. Nat Neurosci 3:1199–1204.
Rissman J, Gazzaley A, D'Esposito M (2004) Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23:752–763.
Rogers TT, Ralph MAL, Hodges JR, Patterson K (2004) Natural selection: the impact of semantic impairment on lexical and object decision. Cogn Neuropsychol 21:331–352.
Saygin ZM, Osher DE, Norton ES, Youssoufian DA, Beach SD, Feather J, Gaab N, Gabrieli JDE, Kanwisher N (2016) Connectivity precedes function in the development of the visual word form area. Nat Neurosci 19: 1250–1255.
Saylik R, Williams AL, Murphy RA, Szameitat AJ (2022) Characterising the unity and diversity of executive functions in a within-subject fMRI study. Sci Rep 12:8182.
Schlochtermeier LH, Kuchinke L, Pehrs C, Urton K, Kappelhoff H, Jacobs AM (2013) Emotional picture and word processing: an fMRI study on effects of stimulus complexity. PLoS One 8:e55619.
Seghier ML, Fagan E, Price CJ (2010) Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. J Neurosci 30:16809–16817.
Seghier ML (2012) The angular gyrus: multiple functions and multiple subdivisions. Neuroscientist 19:43–61.
Sezer I, Pizzagalli DA, Sacchet MD (2022) Resting-state fMRI functional connectivity and mindfulness in clinical and non-clinical contexts: a review and synthesis. Neurosci Biobehav Rev 135:104583.
Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011) Commonality of neural representations of words and pictures. Neuroimage 54:2418–2425.
Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556). arXiv. http://arxiv.org/abs/1409.1556
Sliwinska MW, James A, Devlin JT (2015) Inferior parietal lobule contributions to visual word recognition. J Cogn Neurosci 27:593–604.
Storrs KR, Kietzmann TC, Walther A, Mehrer J, Kriegeskorte N (2021) Diverse deep neural networks all predict human inferior temporal cortex well, after training and fitting. J Cogn Neurosci 33:2044–2064.
Szwed M, Dehaene S, Kleinschmidt A, Eger E, Valabrègue R, Amadon A, Cohen L (2011) Specialization for written words over objects in the visual cortex. Neuroimage 56:330–344.
Tan LH, Perfetti CA (1999) Phonological activation in visual identification of Chinese two-character words. J Exp Psychol Learn Mem Cogn 25: 382–393.
Taylor JSH, Rastle K, Davis MH (2013) Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. Psychol Bull 139:766–791.
Thomas Yeo BT, et al. (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106:1125–1165.
Tsai J-L, Lee C-Y, Tzeng OJL, Hung DL, Yen N-S (2004) Use of phonological codes for Chinese characters: evidence from processing of parafoveal preview when reading sentences. Brain Lang 91:235–244.
Twomey T, Kawabata Duncan KJ, Price CJ, Devlin JT (2011) Top-down modulation of ventral occipito-temporal responses during visual word recognition. Neuroimage 55:1242–1251.
Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RSJ (1996) Functional anatomy of a common semantic system for words and pictures. Nature 383:254–256.
Van Den Heuvel MP, Sporns O (2013) Network hubs in the human brain. Trends Cogn Sci 17:683–696.
Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L (2007) Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron 55:143–156.
von Seth J, Nicholls VI, Tyler LK, Clarke A (2023) Recurrent connectivity supports higher-level visual and semantic object representations in the brain. Commun Biol 6:1–15.
White AL, Kay KN, Tang KA, Yeatman JD (2023) Engaging in word recognition elicits highly specific modulations in visual cortex. Curr Biol 33: 1308–1320.e5.
Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E (2011) The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. Cereb Cortex 21:1066–1075.
Wu C-Y, Ho M-HR, Chen S-HA (2012) A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. Neuroimage 63:381–391.
Xue G, Chen C, Jin Z, Dong Q (2006) Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. Neuroimage 31:1315–1326.
Zhao W, Li Y, Du Y (2021) TMS reveals dynamic interaction between inferior frontal gyrus and posterior middle temporal gyrus in gesture-speech semantic integration. J Neurosci 41:10356–10364.
Zheng Y, Zhang J, Yang Y, Xu M (2025) Neural representation of sensorimotor features in language-motor areas during auditory and visual perception. Commun Biol 8:1–15.
Zhou X, Zhang Z, Yu L, Fan B, Wang M, Jiang B, Su Y, Li P, Zheng J (2021) Disturbance of functional and effective connectivity of the salience network involved in attention deficits in right temporal lobe epilepsy. Epilepsy Behav 124:108308.